REMARKS

Claims 1-37 are pending in this application. Claims 5, 12-17, and 22-31 are withdrawn from consideration. Claims 1-4, 6-11, 18-21, and 32-37 have been examined. Claims 1, 2, 4, 6-9, 18-21, 32, and 34-37 stand rejected. Claims 1, 19, 32 and 37 have been amended. Reconsideration and allowance of Claims 1-4, 6-11, 18-21, and 32-37 is respectfully requested.

Interview Summary

Applicant and applicant's attorney thank the Examiner for granting an interview on September 8, 2005, and for his helpful comments and suggestions during the interview. The participants in the interview were Examiner Dr. Ardin H. Marschel, Examiner Anna Subinski (present as an observer), the applicant and inventor Dr. Erik Gunther (in person), and applicant's attorney Tineka Quinton (by telephone). Applicant provided a document with figures and text for the purpose of clarifying the differences between the Daniel reference and the present invention (included herewith as Exhibit A). Applicant explained that the text presented in Exhibit A consists of direct quotations from the specification of the Daniel et al. reference (U.S. Patent No. 6,368,794), and direct quotations from the specification of the instant application.

Applicant pointed out that the Daniel reference discloses genes that are identified as drugs by their differential expression in cancerous or non-cancerous biological samples. Applicant's attorney noted that the Daniel reference includes a claim directed to a composition comprising an identified cDNA in a suitable carrier (Claim 9), which, for patentability purposes must be adequately supported by a specific, credible utility as a therapeutic.

In contrast, applicant pointed out that the present invention is directed to "the screening of compounds with no previously known pharmacological action, in order to identify drug candidates by virtue of molecular expression profile." Page 14, lines 19-22. The applicant read the following language from the specification: "[T]he method of the invention does not include

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

the screening of compounds with previous biochemically-derived evidence of specific utility (i.e. established drug candidates), in order to identify toxicological molecular expression profiles." Page 14, lines 24-27. "Nor does the method of the invention include the screening of established drug candidates, in order to more fully characterize the utility for which they have already been indicated." Page 14, lines 28-30.

The claim limitation "analyte of previously uncharacterized specific pharmacological activity" in independent Claims 1 and 32 was discussed. An agreement was reached between applicant, applicant's attorney and the Examiner that the phrase "analyte of previously uncharacterized specific pharmacological activity with respect to the parameter by which the first and second samples are known to differ," would be helpful to clarify the scope of the invention.

The Examiner stated that he would remove the Daniel reference with respect to independent Claims 1 and 32, provided that the quotations from the present application as provided in Exhibit A are accurate.

The Rejection of Claims 1, 2, 6-9, 18-21, 32 and 35-37 Under 35 U.S.C. § 103(a) over Daniel et al., U.S. Patent No. 6,368,794

The Examiner has rejected Claims 1, 2, 6-9, 18-21, 32 and 35-37 under 35 U.S.C. § 103(a) as being obvious over U.S. Patent No. 6,368,794 (Daniel et al.). Independent Claim 1, from which Claims 2, 6-9, and 18-21 depend, has been amended to clarify the scope of the invention and recites as amended "at least one analyte of previously uncharacterized specific pharmacological activity with respect to the parameter by which the first and second samples are known to differ." Independent Claim 32, from which Claims 35-37 depend, has been amended to clarify the scope of the invention and recites as amended "at least one analyte of previously uncharacterized specific pharmacological activity with respect to the drug treatment."

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

Applicants submit that the claims, as amended, are not obvious in view of Daniel et al. As mentioned in the interview summary, the figures and text included herewith as Exhibit A describe that Daniel discloses genes that are identified as drugs by their differential expression in cancerous or non-cancerous biological samples. Daniel does not teach or suggest a method of identifying analytes that induce a third expression profile that is more similar to a first expression profile than is a second expression profile, comprising: (a) determining a first expression profile of a set of representative molecules in a first biological sample; (b) determining a second expression profile of the set of molecules in a second biological sample, wherein the second biological sample differs from the first biological sample by a known parameter; (c) determining a third expression profile of the set of molecules in the second biological sample after treatment of the second biological sample with at least one analyte of previously uncharacterized specific pharmacological activity with respect to the parameter by which the first and second samples are known to differ; and (d) comparing the third expression profile with the first and second expression profiles to identify one or more analytes that induces a third expression profile that is more similar to the first expression profile than is the second expression profile, as now claimed.

For the reasons stated above and the reasons already of record, applicant submits that Daniel et al. does not teach, suggest, or provide any motivation to arrive at the claimed invention, and respectfully requests withdrawal of this ground of rejection.

The Applicant's Published Scientific Articles as Evidence of the Successful Practice of the Invention

The applicant provides the following scientific articles as evidence of the successful practice of the invention: E.C. Gunther et al., *Pharmacogenomics Journal*, 1470-1478 (2005) (included herewith as Exhibit B), E.C. Gunther et al., *PNAS* 100: 9608-9613 (2003) (included herewith as Exhibit C). The applicant also provides a review article by S. E. Levy published in

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

Science's STKE, in October 2003 (included herewith as Exhibit D) which comments on the revolutionary nature of the present invention.

Examination of Non-Elected Species

In the election requirement mailed September 22, 2003, Claims 1-4, 6-8, and 18-21 were found to be generic. The species "polynucleic acid microarrays" was elected for initial examination. Applicant respectfully submits that the pending claims are allowable with respect to polynucleic acid microarrays and request rejoinder of the claims that have been withdrawn.

CONCLUSION

In view of the foregoing amendments and remarks, the application is believed to be in condition for allowance. If any issues remain that can be expeditiously addressed in a telephone interview, the Examiner is encouraged to telephone applicant's attorney at 206.695.1655.

Respectfully submitted,

CHRISTENSEN O'CONNOR
JOHNSON KINDNESS[PLLC]

Tineka J. Quinton

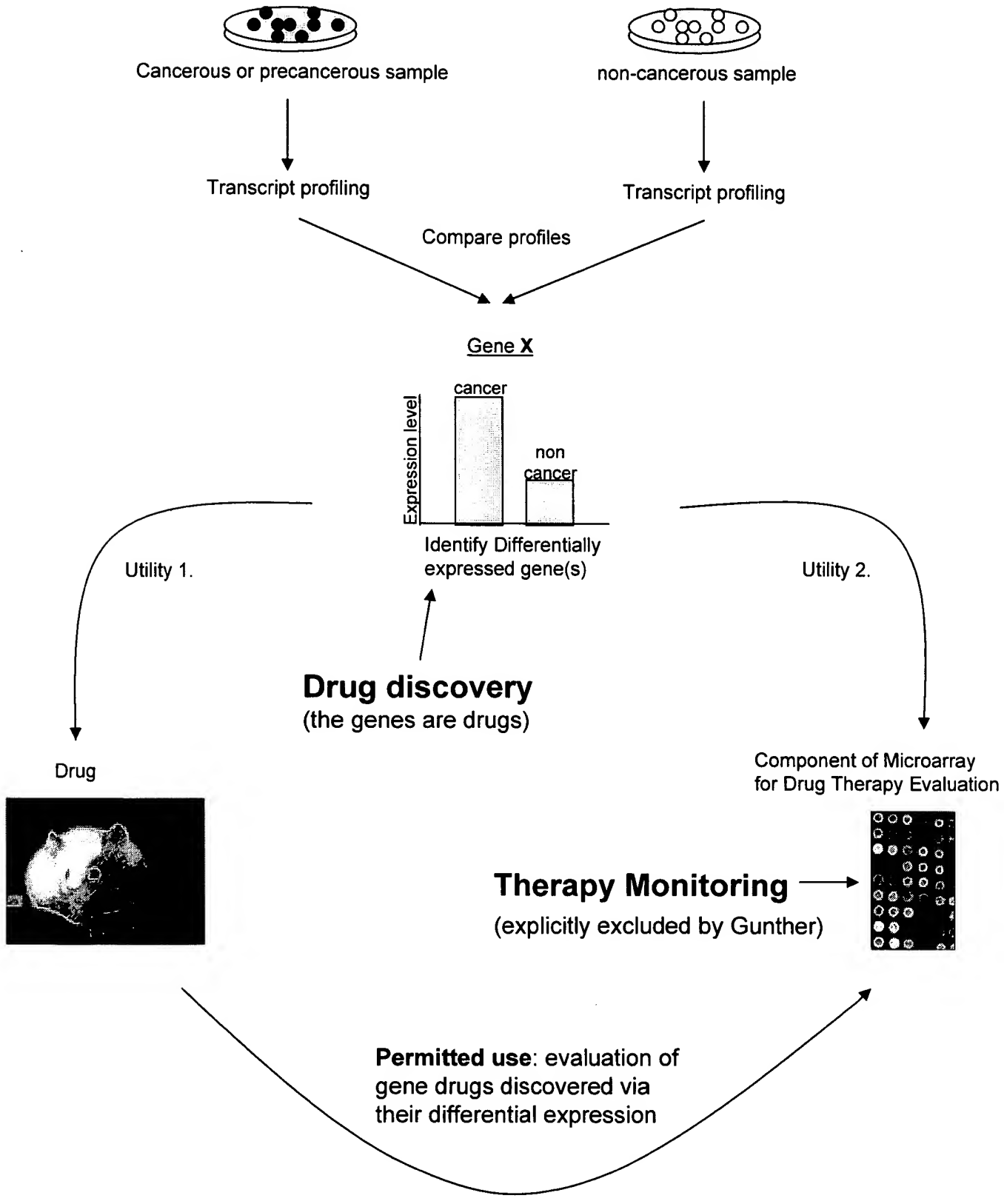
Tineka J. Quinton
Registration No. 53,496
Direct Dial No. 206.695.1655

I hereby certify that this correspondence is being deposited with the U.S. Postal Service in a sealed envelope as first class mail with postage thereon fully prepaid and addressed to Mail Stop RCE, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the below date.

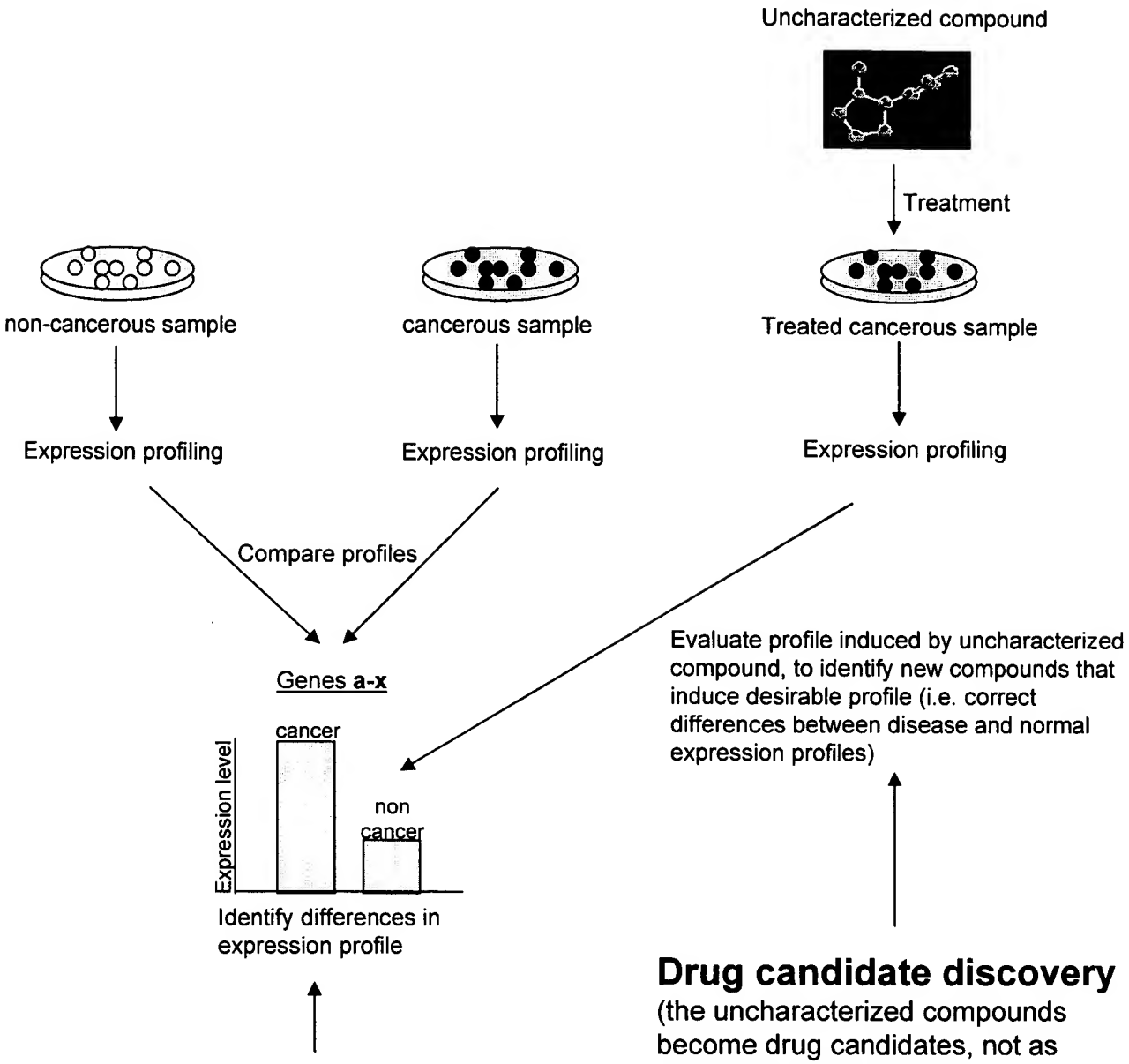
Date: September 9, 2005 Pamela M Tucker

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS[PLLC]
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

Daniel et al. summary (one aspect)



Gunther summary (one aspect)



Not Drug Discovery
(the differentially regulated genes are not being identified as drugs, as in Daniel et al.)

Drug candidate discovery
(the uncharacterized compounds become drug candidates, not as described in Daniel et al.)

Daniel et al. Summary – annotations for diagram

1. Drug discovery

"...first polynucleotide sequences are selected by a first method comprising... comparing transcript profiles to detect a plurality of genes that are differentially expressed in either non-cancerous, precancerous and cancerous biological samples..." (column 1, line 55-62).
"...each of said polynucleotides comprises at least a fragment of a gene implicated in the regulation of cell proliferation" (column 1, line 42-45).
"... the invention entails a pharmaceutical composition comprising a polynucleotide or polypeptide in conjunction with a suitable pharmaceutical carrier and a method for treating or preventing a disease or condition associated with the altered expression of genes that regulate cell proliferation comprising administering to a subject in need such a composition an amount effective for treating or preventing said disease." (Column 2-3, line 65-5).
"The phrase 'genes that regulate cell proliferation' refers to genes whose altered expression results in a cancerous or a precancerous stage in a biological sample." (column 4, line 38-40).

Synopsis: Genes are identified as drugs by their differential expression.

2. Monitoring established treatments

"The composition is particularly useful as hybridizable elements in a microarray for monitoring the expression of a plurality of sample polynucleotides implicated in theregulatio of cell proliferation. The microarray can be used, for example, in the prognosis, diagnosis and treatment of a cancer or other disease relating to the altered expression of genes involved in cell proliferation." (column 2, line 25-31).
"...the present invention provides a method for diagnosing, prognosing, or monitoring the treatment of a disease." (column 2, line 43-45).
"...a transcript profile can show differences occurring... between treated and untreated tissues, such as prostate tumor and irradiated prostate tumor." (column 6, line 17-21).
"The sequences of the invention may be used in diagnosis, prognosis, treatment, prevention, and evaluation of therapies for diseases associated with cell proliferation." (column 10, line 5-7).
"Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient. Once the presence of a disease is established and a treatment protocol initiated, hybridization or amplification assays can be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in a healthy subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months." (column 10-11, line 64-7).

Synopsis: The sequences can be used to monitor/evaluate the course of existing therapies.

Gunther Summary – annotations for diagram

A method for identifying analytes that induce a third expression profile that is more similar to a first expression profile than is a second expression profile, comprising,

a) determining a first expression profile of a set of representative molecules in a first biological sample;
b) determining a second expression profile of the set of molecules in a second biological sample, wherein the second biological sample differs from the first biological sample by a known parameter;
c) determining a third expression profile of the set of molecules in the second biological sample after treatment of the second biological sample with at least one analyte of previously uncharacterized specific pharmacological activity, and
d) comparing the third expression profile with the first and second expression profiles to identify one or more analytes that induces a third expression profile that is more similar to the first expression profile than is the second expression profile. (Claim 1)

"previously uncharacterized chemicals that exhibit a specific biochemical activity revealed by the screen are reclassified as 'candidate drugs, also known as 'hits', 'drug candidates', and 'drug leads'". (Page 2, line 19-22).

"As used herein, an 'analyte' refers to a compound that is being tested for its impact on a particular expression profile when exposed to a biological sample." (page 7, line 18-19)

"One of the strategies in biomedical research today is to elucidate the underlying genetic architecture involved in complex traits, so that pharmaceuticals may be generated that specifically impinge on the relevant components of that architecture. This goal has been pursued with the use of microarrays and other techniques of molecular expression analysis in order to identify specific genes and pathways underlying diseases and drug response. These genes and pathways then become the focus of functional study and the targets for pharmaceutical development. The approach of this invention differs from that strategy in practice and philosophy. The present invention utilizes the molecular expression profile as a quantifiable symptom of a pathology, rather than as a means of understanding the underlying root cause so that the root cause may be addressed in a directed fashion with the creation of drugs targeted to specific disease mechanisms." (page 13-14 , line 27-7).

"Another difference between the method of this invention and past practices is that the method encompasses the screening of compounds with no previously known pharmacological action, in order to identify drug candidates by virtue of molecular expression profile." (Page 14, line 19-22). "The method of the invention does not include the screening of compounds with previous biochemically-derived derived evidence of specific utility (i.e. established drug candidates), in order to identify toxicological molecular expression profiles. Nor does the method of the invention include the screening of established drug candidates, in order to more fully characterize the utility for which they have already been indicated." (Page 14, line24-30).

The Pharmacogenomics Journal (2005), 1–9

www.nature.com/tpj

ORIGINAL ARTICLE

A quantitative genomic expression analysis platform for multiplexed *in vitro* prediction of drug action

EC Gunther[1]
DJ Stone[1,2]
JM Rothberg[1] and
RW Gerwien[1]

[1]*CuraGen Corporation, Branford, CT, USA*

Correspondence:
Dr EC Gunther, Yale University School of Medicine, Department of Neurology, YPI CP314, 100 John Murphy Dr., New Haven, CT 06513, USA. Tel: +1 203 705 5030 Fax: 1 203 785 5098 E-mail:erik.gunther@yale.edu

[2]Current address: Rosetta Inpharmatics, Seattle, WA, USA

Received 6 August 2004
Revised 12 October 2004
Accepted 3 December 2004

ABSTRACT

Genomic expression signatures provide high-content biomarkers of cellular physiology, including the diverse responses to therapeutic drugs. To recognize these signatures, we devised a method of biomarker evaluation called 'sampling over gene space' (SOGS) that imparts superior predictive performance to existing supervised classification algorithms. Applied to microarray data from drug-treated human cortical neuron 1A cell cultures, this method predicts whether individual compounds possess anticonvulsant, antihypertensive, cyclooxygenase inhibitor, or opioid action. Thus, stable cell lines can be suitable for expression signature-based screening of a diverse range of activities. A SOGS-based system also discriminates physiologically active from inactive compounds, identifies drugs with off-target side effects, and incorporates a quantitative method for assigning confidence to individual predictions that, at its most stringent, approaches 100% accuracy. The capacity to resolve multiple distinct drug activities while simultaneously discriminating inactive and potential false-positive compounds in a cell line presents a unified framework for streamlined chemical genomic drug discovery.
The Pharmacogenomics Journal advance online publication, 22 February 2005; doi:10.1038/sj.tpj.6500300

Keywords: efficacy profiling; predictive efficacy; drug discovery; fingerprinting; supervised classification

INTRODUCTION

Genomic expression signatures provide a generic indicator of diverse cellular physiologies.[1–8] Recently, expression signatures have been used to identify clinically relevant cellular physiologies that result from drug treatments *in vitro*.[9–11] A pattern recognition-based approach has potential utility for multiplexed *in vitro* drug discovery, as gene expression signatures constitute a comprehensive assay for disparate drug activities, and can predict clinical drug efficacy directly from cellular physiology, even when the specific drug target is unknown.[11] However, previous chemical genomic studies with mammalian cells have resolved only a single type of drug activity[9,10] or employed primary cell co-culture systems apt to exhibit batch-to-batch inconsistency.[11] For the potential of this discipline to be met, it will be necessary to establish a simple cellular platform capable of evaluating multiple distinct clinical efficacies at once for each compound screened. It will also be necessary to develop a practical, quantitative means to prioritize the most promising hits for development as therapeutics.

Here, we describe an efficacy-profiling platform, based on the human cortical neuron 1A (HCN-1A) cell line, capable of predicting multiple different drug class activities with good precision. Using microarray data from cells treated with multiple examples of nine distinct drug classes, we develop a feature assessment algorithm called 'sampling over gene space' (SOGS), that works well with small sample numbers, does not make assumptions about the behavior of individual features, and yields a convenient metric for ranking prediction strength based on a vote-counting algorithm. We employ SOGS to distinguish drug classes that are active or inactive in our cell culture system. Then, we use SOGS to build multiplexed statistical models of drug action with better predictive accuracy than models built with standard supervised classification methods. Finally, by applying the SOGS-based strength metric to these predictions, we establish a ranking system for assigning confidence to chemical genomic drug utility predictions.

RESULTS

Sampling Over Gene Space

Supervised classification methods learn from example data sets of known class to identify data features (biomarkers) that can classify unfamiliar data sets. High-dimensional microarray data, however, can lead to models that are 'overfit' on distracting features unrelated to the phenomenon being modeled. Algorithms that employ stochastic feature evaluation, such as Random Forest (RF), are resistant to overfit, but the recursive partitioning aspect of RF can lead to high model variance and weak predictions. Other classification methods such as linear discriminant analysis (LDA) and support vector machine (SVM) typically make stronger predictions, but lack the resistance to overfit, imparted by the stochastic aspects of RF. To incorporate the benefits of stochastic feature evaluation into LDA and SVM, we developed a procedure called SOGS that randomly samples a set of features for each episode of model construction by LDA or SVM. Multiple iterations of model construction yield a compendium of class assignments for the unknown. The class receiving the greatest overall number of classifications 'wins' the final prediction. Such a combination of stochastic feature evaluation with the stable LDA and SVM modeling methods minimizes overfit, while increasing prediction strength.

Active *vs* Inactive Drug Class Discrimination

Undifferentiated cultures of the HCN-1A human cortical neuron cell line were treated with 83 different drugs comprising nine defined pharmaceutical classes: antianxiety, anticonvulsant, antidepressant, antihypertensive, antipsychotic, cyclooxygenase (COX) inhibitor, matrix metalloprotease (MMP) inhibitor, opioid receptor agonist, and statin. Each of these drug classes was generally characterized for activity *vs* inactivity in this culture system by comparing microarray data from each pairwise with the control vehicle-treated class, using LDA. LDA was conducted either with ANOVA-filtered data, or by incorporating SOGS to analyze unfiltered data. Drug classes were scored as either active or inactive, as indicated by prediction *P*-value. LDA performed with ANOVA yielded $P>0.05$ for all drug classes and was thus unable to discriminate any drug class treatment from control treatment (Table 1). In contrast, LDA performed with SOGS identified five pharmaceutical classes (antianxiety, anticonvulsant, antihypertensive, COX inhibitor, and opioid) that exhibited genomic expression distinct from controls with $P\leqslant 0.05$, and thus deemed active in our culture system; four classes (antidepressant, antipsychotic, MMP inhibitor, and statin) were statistically indistinguishable from controls with $0.68\leqslant P\leqslant 1.0$ (Table 1), and thus deemed inactive.

Table 1 General activity evaluation

SOGS	*SOGS*		*ANOVA*	
Drug class	*Accuracy*	*P-value*	*Accuracy*	*P-value*
Antianxiety	84.6	0.03	61.5	0.38
Anticonvulsant	80.0	0.03	66.6	0.23
Antidepressant	64.0	1.00	76.0	0.11
Antihypertensive	85.7	0.01	71.4	0.13
Antipsychotic	65.2	0.68	60.9	0.62
COX inhibitor	86.7	0.01	66.6	0.23
MMP2 inhibitor	46.7	0.82	73.3	0.10
Opioid	78.6	0.05	64.3	0.28
Statin	25.0	0.99	16.7	1.0

LDA was conducted using either SOGS or an ANOVA filter in a comparison of each drug class *vs* control class. Accuracy, percent of drug-treated and control samples correctly predicted in leave-one-out supervised pairwise classification. *P*-value; likelihood of random class separation. Shaded drug classes are active in this culture (meet the $P\leqslant 0.05$ cutoff for presence of drug activity). LDA using SOGS, but not ANOVA, distinctly identifies five active drug classes.

Prediction of Drug Efficacy by Supervised Classification

The value of SOGS to multiplexed drug efficacy prediction was investigated. LDA and SVM were each conducted with or without the use of SOGS, to analyze data from the five drug classes that exhibit activity in HCN1A cell culture (antianxiety, anticonvulsant, antihypertensive, COX inhibitor, and opioid). These methods were conducted in a 'leave-one-out' fashion, where every sample is treated once as an unknown in an independent episode of model construction, allowing an objective assessment of predictive accuracy with unfamiliar compounds. LDA and SVM exhibited 39.5 and 44.2% predictive accuracy, respectively, when used to directly analyze microarray data. By comparison, when SOGS was incorporated, the predictive accuracy of LDA and SVM rose to 67.4% (Figure 1) and 58.1%, respectively. When RF, which already contains a stochastic element, was conducted using the same data, a predictive accuracy of 58.1% was achieved. Thus, by introducing a stochastic element to LDA and SVM, SOGS imparted predictive accuracy equal to or better than the tree-based RF supervised classification technique.

a

Compound	True Class	Pred Class	CI
alprazolam	AA	AA	0.33
zopiclone	AA	AA	0.29
flurazepam	AA	AC	0.51
triazolam	AA	AC	0.39
zolpidem	AA	OP	0.35
diazepam	AA	AC	0.33
riluzole	AC	AC	0.55
carbamazepine	AC	AC	0.53
ethosuximide	AC	AC	0.42
felbamate	AC	AC	0.42
vigabatrin	AC	AC	0.39
gabapentin	AC	AC	0.38
phenytoin	AC	AA	0.29
valproate	AC	AH	0.29
verapamil	AH	AH	0.86
nimodipine	AH	AH	0.71
nitrendipine	AH	AH	0.67
niguldipine	AH	AH	0.35
nifedipine	AH	AA	0.30
YS 035	AH	AC	0.30
diltiazem	AH	AC	0.27
DMSO 1	CON	CON	0.73
DMSO 4	CON	CON	0.67
DMSO 3	CON	CON	0.64
DMSO 5	CON	CON	0.64
DMSO 2	CON	CON	0.57
DMSO 6	CON	CON	0.39
DMSO 7	CON	AA	0.41
aspirin	COX	COX	0.82
dup.697	COX	COX	0.68
flurbiprofen	COX	COX	0.67
NS398	COX	COX	0.62
ibuprofen	COX	COX	0.48
sulindac	COX	COX	0.38
indomethacin	COX	OP	0.42
naproxen	COX	CON	0.40
U62066	OP	OP	0.50
DAMGO	OP	OP	0.40
U50488	OP	OP	0.38
DALDA	OP	OP	0.32
enkephalin	OP	OP	0.29
endomorphin	OP	COX	0.41
dynorphin.A	OP	AA	0.33
Percent correct	100	67.4	

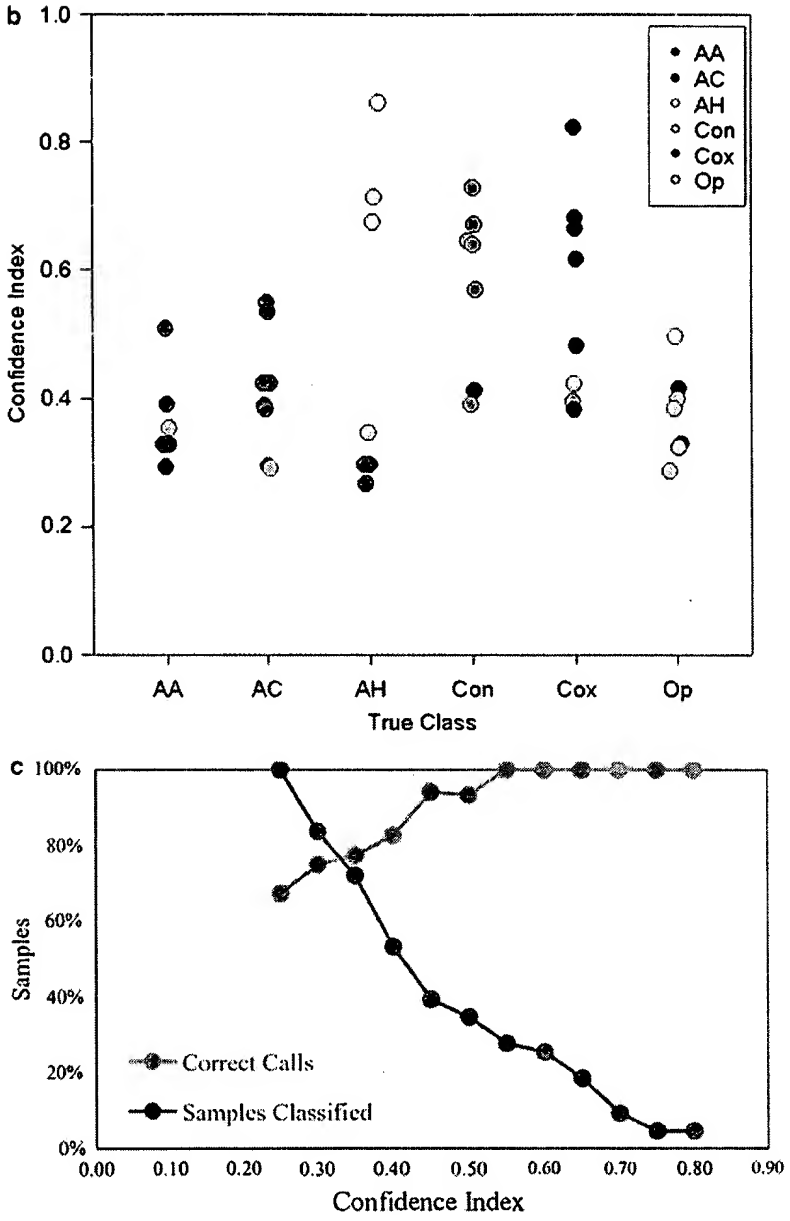


Figure 1 (a) Drug class predictions and CI scores: compounds active in culture. True Class, known therapeutic efficacy; Pred Class, drug class prediction by SOGS-based LDA for the expression profiles elicited by the drugs listed on the left; CI, SOGS-derived CI measure of prediction strength; shaded designations, correct predictions; AA, antianxiety; AC, anticonvulsant; AH, antihypertensive; Con, control; Cox, cyclooxygenase inhibitor, Op, opioid. (b) Confidence ranking by class: active drug predictions. True Class, known therapeutic efficacy; colored dots, predicted efficacy. (c) CI *vs* prediction accuracy. Correct calls, percent correct prediction of samples with a CI score $\geqslant$ the corresponding value. Samples classified, percent of total samples with a CI score $\geqslant$ the corresponding value.

Graphical Class Separation

SVM, LDA, and RF each establish hyperboundaries between sample types in *n*-dimensional gene space; biomarkers given the greatest weight provide optimal class separation. Three-dimensional graphical depiction of the first three linear discriminants of the 12 most heavily weighted biomarkers identified by LDA (without SOGS), for example (Figure 2b), illustrates the drug efficacy class separation provided by this gene assemblage (Figure 2a).

Confidence Index

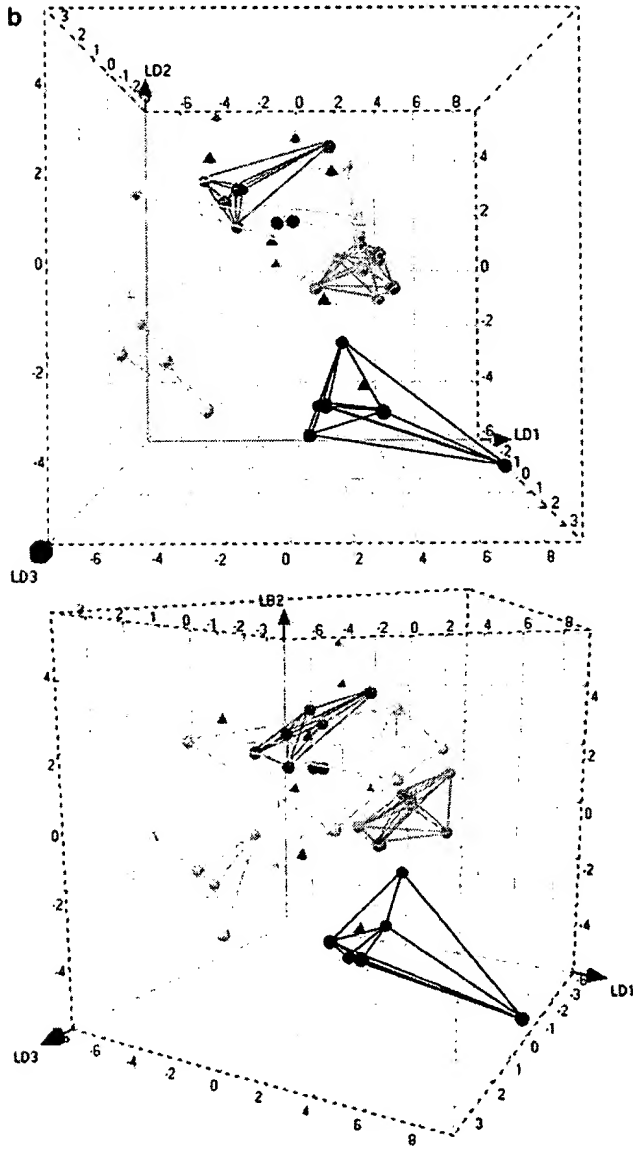
Supervised classification methods make predictions based on the number of votes per class received by a given sample; the category receiving the greatest number of votes wins the

prediction. In this winner-take-all scheme, predictions based on votes won by slight or wide margins are indistinguishable. To utilize the information contained in the vote margins for each sample, we devised a confidence index (CI) that indicates the winning vote proportion from the SOGS-based LDA models (Figure 1a). The CI provides a continuous output from zero-to-one, indicating the strength of association of a sample with its predicted class. Comparison of this CI with the predictions indicated that the magnitude of the CI scales closely with the accuracy of the predictions (Figure 1b, c): 94.1% of predictions above a CI of 0.45 are accurate, while 50.0% of predictions below a CI of 0.45 are incorrect. Thus, the CI derived from SOGS-based modeling is an effective general measure of prediction quality.

a

Marker Identity	Marker Count
PC - pyruvate carboxylase	28
FLJ23342 - hypothetical protein	27
CBS - cystathionine-beta-synthase	24
PDE8A - phosphodiesterase 8A	21
YARS - tyrosyl-tRNA synthetase	20
BIRC1 - baculoviral IAP repeat-containing 1	20
RAP2A - ras-related GTPase	20
TULP3- tubby like protein 3	18
BEXL1- brain expressed X-linked-like 1	17
STC2 - stanniocalcin 2	17
CHST12 - carbohydrate sulfotransferase 12	17
REC8L1 - REC8-like 1	17

b



Classification and Confidence Ranking of Inactive Compounds

The finding that SOGS-based LDA effectively discriminated untreated controls from active compounds (Figure 1) suggested that inactive compounds might also be effectively classified. To investigate this capacity, we presented the expression data from the drugs found to be inactive in HCN1A cell culture (antidepressant, antipsychotic, MMP inhibitor, and statin, Table 1) to a SOGS-based LDA model constructed using data from active drug treatments. As expected for inactive compounds, a large percentage of these samples (51.1%) were classified as controls (Figure 3). To characterize the remaining presumptive false positives, a CI value was derived for each prediction and ranked by magnitude. Similarly to the active compound predictions, a high CI value corresponded to accurate 'control' prediction of inactive classes, while a low CI value corresponded to activity mispredictions: 80.8% of predictions above a CI of 0.45 were accurate, while 85.7% of predictions below a CI of 0.45 were incorrect. Thus, the CI serves as a means of prioritizing both correctly predicted efficacies and correctly predicted inactive compounds.

Classification and Confidence Ranking of Out-group Activities

In addition to accurately assessing drug activity and inactivity, the ideal drug assay would exclude active compounds that elicit physiology distinct from the desired activities (ie, drugs that induce side effects). Such prospective false positives were simulated by omitting each entire

Figure 2 (a) Top 12 LDA biomarkers. LDA performed with data from the five active drug classes and controls identified 12 markers with counts of 17 or greater. Marker count, number of times out of the total 43 model building episodes a particular gene was selected as a marker. (b) Spatial class separation by top 12 LDA biomarkers. Axes represent the first three linear discriminants of the expression levels of the 12 genes from Figure 3a. The graph is shown twice, perpendicular to the XY plane (left) and from 20° rotation around the *y*-axis (right). Red, antianxiety; dark blue, anticonvulsant; yellow, antihypertensive; green, DMSO control; black, COX inhibitor; turquoise, opioid; circles, correctly predicted treatments; triangles, misclassified treatments. Lines connecting correctly predicted treatments delineate the hypervolume occupied by each accurately defined sample class.

a

Compound	True class (inactive)	Pred class	CI
trazodone	AP	CON	0.79
444244	MI	CON	0.78
tranylcypromine	AD	CON	0.72
pravastatin	ST	CON	0.69
desipramine	AD	CON	0.69
amoxapine	AD	CON	0.69
444241	MI	CON	0.67
444240	MI	CON	0.65
paroxetine	AD	AC	0.61
simvastatin	ST	CON	0.59
fluvastatin	ST	CON	0.59
nortryptyline	AD	CON	0.58
loxapine	AP	CON	0.58
444264	MI	CON	0.57
amitriptyline	AD	AC	0.57
citalopram	AD	AH	0.57
bupropion	AP	CON	0.56
444242	MI	CON	0.56
thioridazine	AP	AC	0.56
clomipramine	AD	AC	0.52
clozapine	AP	CON	0.51
444249	MI	CON	0.50
mevastatin	ST	CON	0.50
trimipramine	AD	CON	0.49
lovastatin	ST	CON	0.49
phenelzine	AD	CON	0.46
chlorprothixene	AP	AC	0.44
thiothixene	AP	AC	0.43
fluphenazine	AP	CON	0.42
maprotiline	AD	COX	0.41
triflupromazine	AP	OP	0.41
perphenazine	AP	AC	0.38
protriptyline	AD	AC	0.38
fluoxetine	AD	AC	0.38
444251	MI	OP	0.37
droperidol	AP	CON	0.37
risperidone	AP	AC	0.36
doxepin	AD	COX	0.36
imipramine	AD	AC	0.35
haloperidol	AP	AC	0.35
chlorpromazine	AP	AC	0.33
364205	MI	COX	0.31
iproniazid	AD	AA	0.29
pimozide	AP	AC	0.29
sertraline	AD	OP	0.28
trifluoperazine	AP	OP	0.26
fluvoxamine	AD	CON	0.23
Percent correct		51.1	

b

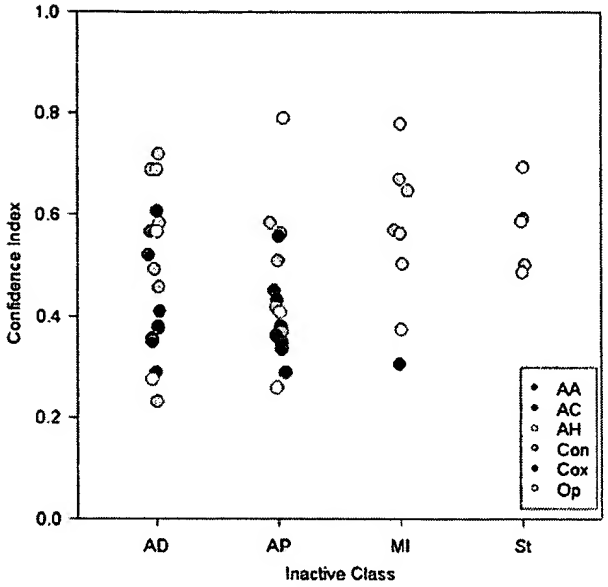


Figure 3 (a) Drug class predictions and CI scores: compounds inactive in culture. Inactive drug classes predicted by model built from active drug classes. True Class, known clinical efficacy; Pred Class, predicted efficacy (shaded designations, correct predictions); ST, statin; AP, antipsychotic; AD, antidepressant; MI, MMP inhibitor; AA, antianxiety; AC, anticonvulsant; AH, antihypertensive; Con, control; Cox, cyclooxygenase inhibitor, Op, opioid. Percent correct, proportion predicted as controls. (b) Confidence ranking by class: inactive drug predictions. Inactive class, true drug utility; AD, antidepressant; AP, antipsychotic; MI, MMP inhibitor; St, statin; colored dots, predicted activity.

active drug class from a SOGS-based LDA model built with data from the remaining four active classes and controls. Data from the members of the omitted class were presented to the model for prediction, and the CI for each prediction was calculated (Figure 4). By definition, these out-group drug predictions are inherently incorrect; as such, they would be expected to have a low overall CI compared to correctly predicted active compounds. Accordingly, the average out-group CI value of 0.35 ± 0.012 was indistinguishable from the average CI of 0.36 ± 0.018 for incorrect

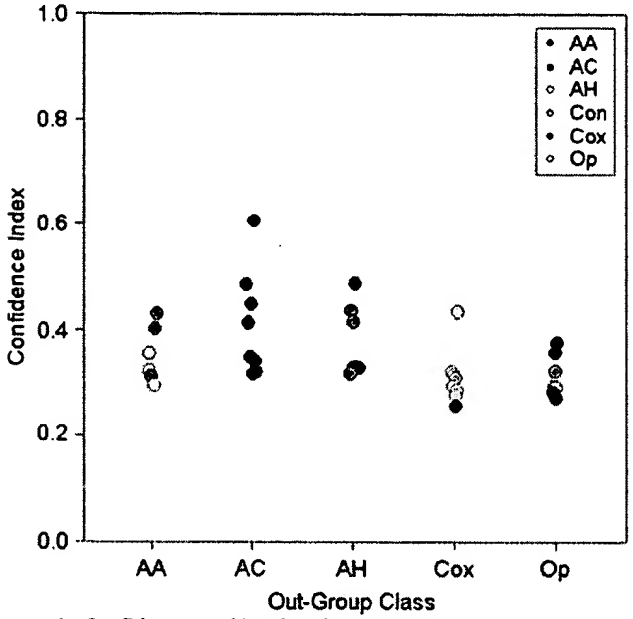


Figure 4 Confidence ranking by class: out-group activity mispredictions. Entire active drug classes were serially withheld as out-groups from models built with remaining active drug data, and the CI of each inherent misprediction was calculated. Out-group class, true utility; colored dots, class of misprediction. AA, antianxiety; AC, anticonvulsant; AH, antihypertensive; Con, control; Cox, cyclooxygenase inhibitor, Op, opioid.

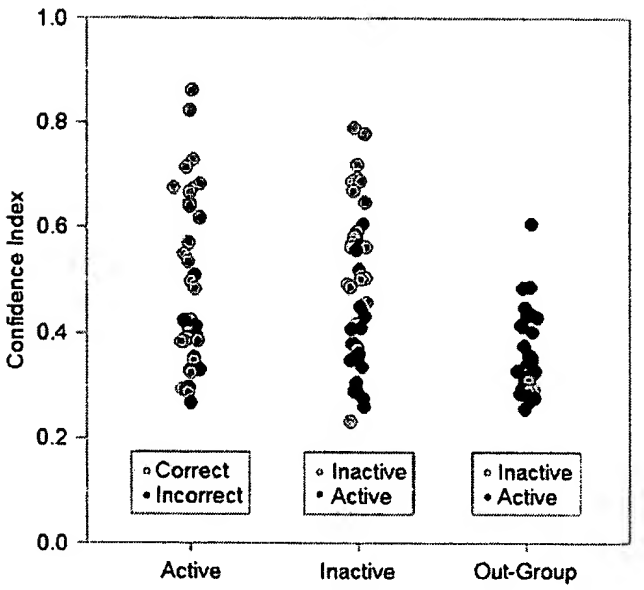


Figure 5 Correlation of CI with prediction accuracy: summary of active, inactive and off-target class predictions. Active category green and red dots, correctly and incorrectly predicted active-class drugs, respectively; Inactive category green and red dots, inactive-class drugs correctly and incorrectly predicted as being inactive or active, respectively; out-group green and red dots, mispredictions of off-target active drug efficacy.

activity predictions (Figure 1), and substantially lower than the average CI of 0.52 ± 0.031 for accurate activity predictions (Figure 1). Thus, the CI can be used as a universal ranking measure of true drug activity, drug inactivity, and off-target (nonspecific) activity (Figure 5), *in vitro*.

DISCUSSION

Previously, we found that primary cultures derived from brain tissue exhibited genomic expression profiles that allowed prediction of antidepressant, antipsychotic, and opioid drug efficacy.[11] However, the use of primary cultures, while appealing physiologically, presents inherent experiment-to-experiment consistency issues that must be addressed in order to establish a stable screen. Cell line monoculture provides a simpler system that is likely to be more amenable to reproducible chemical genomics, but the ability of immortalized cell lines to respond to multiple disparate drugs with signature expression profiles has not been explored. Previous chemical genomic studies have also lacked a quantitative means of identifying drugs that have nonspecific effects outside the targeted desirable pharmaceutical effect. The ability to discriminate such nonspecificity early in drug discovery would be a substantial advantage, because many candidate pharmaceuticals fail late in development due to unforeseen side effects. Consequently, we sought to determine whether a cell line-based activity prediction platform could be established as a unified guide for identifying multiple distinct drug utilities, while concurrently distinguishing inactive compounds and active drugs with undesired 'side' effects.

Microarray data present special challenges to existing supervised classification statistical techniques because of the abundance of uninformative features. Data pre-filtering techniques such as ANOVA are typically used to focus on informative features, but can make erroneous distributional assumptions that cause predictive models to be corrupted by noise. Accordingly, the supervised classification method of LDA using an ANOVA filter was unable to distinguish drug classes that were active or inactive in culture when compared pairwise with controls (Table 1). However, SOGS, which makes no *a priori* assumptions about data characteristics, enabled LDA to clearly parse the drug treatments into active and inactive classes (Table 1), even though no data pre-filtering was used to reduce noise. The objective nature of SOGS appears to improve class resolution for situations where sample number is small, data quality is noisy, and many features contribute to class differences—all common aspects of high-dimensional microarray experiments. In multiplexed activity prediction, the incorporation of SOGS into the widely used SVM and LDA methods enabled them to equal or exceed the performance of RF, the stochastic nature of which otherwise renders it superior in this application. Thus, SOGS provides an advantageous technique for contending with the noise that afflicts predictive microarray analyses.

The finding that undifferentiated HCN-1A cells provide a range of diverse genomic responses to drug action indicates that cell lines can be a suitable alternative to primary cell cultures for multiplexed predictive chemical genomics. Our assay effectively predicted the drug classes of anticonvulsant, antihypertensive, COX inhibitor, and opioid, which act, respectively, on diverse cellular targets such as ion channel, enzyme, or G protein-coupled receptor. Typically, biochemical assays for such disparate activities are carried out independently. In contrast, the resolution of these activities by a single cell type using a single type of output, that is, gene expression profile, enables the screening of individual compounds for multiple functions simultaneously. The presence of substantial complexity at the systems biology level in HCN-1A cells, as suggested by expression of over 50% of the genes on our microarray, is likely to provide many additional functional points for pharmacological intervention. Studies in yeast have found that approximately 50% of gene deletions lead to recognizable transcription pattern changes, corresponding to a range of physiological processes.[3] If these studies are an indication, the potential for mammalian cell-based multiplexed drug activity prediction may be substantial. Previously, we have shown that primary neuronal cultures can be used to predict the efficacy of antidepressant and antipsychotic drug classes.[11] The apparent lack of responsiveness of undifferentiated HCN-1A cell cultures to the antidepressant, antipsychotic, statin, and MMP inhibitor drug classes may result from the lack of expression of active targets or components of the signal transduction machinery that mediate some of these activities. Alternatively, this culture system may exhibit a time- or dose-dependent course of physiological response not captured by our treatment regimens. Future predictive accuracy is likely to be improved by optimizations of these treatment parameters as well as utilization of culture systems that possess the appropriate systems biology for particular drug responses of interest.

To enhance the accuracy of activity profiling, we sought to create a system for determining which predictions are likely to be accurate. Again, SOGS provided a reliable means of doing so. Using the SOGS-derived CI, predictions were ranked by strength in a manner that closely parallels prediction accuracy. The broad utility of the CI as an identifier of drug nonspecificity (Figure 4) or inactivity (Figure 3), as well as true drug efficacy (Figure 1), provides the core of a concerted *in vitro* system of drug efficacy prediction. When we modeled each of these aspects of drug action individually, ranking by CI effectively sorted the predictions by accuracy. When these prediction types are considered together as a function of CI (Figure 5), values may be chosen that provide a desired stringency of efficacy prediction accuracy and simultaneous intolerance of false positives (drugs that induce physiologies other than those encompassed by the assay). This capacity is of practical importance in drug development because many candidates are disqualified as therapeutics for possessing side effects or toxicity resulting from nonspecific activity. In current practice, these side effects are usually discovered late in the development process, imposing substantial burden. By identifying side effects at an early stage, this inefficiency can be minimized.

In practice, it will be necessary to establish a balance between the accurate predictions that are captured, and the false positives that are allowed. For example, above a CI threshold of 0.61, 100.0% of target activities and 100% of inactive compounds identified are correct, while 100.0% of off-target activities are excluded (Figure 5). However, this stringent CI boundary has the effect of deprioritizing a large portion of the accurately predicted target activities as well as restricting the classes captured to only the most confidently predicted antihypertensives and COX inhibitors. In order to capture a higher percentage of the accurately predicted true activities and encompass anticonvulsant and opioid predictions, a lower CI value could be chosen with the trade-off of tolerating a higher proportion of false-positive predictions. Above a CI threshold of 0.45, for example, the strongest activity predictions from the four most predictable drug classes are captured with an accuracy of 94.1%, while tolerating 19.2% inaccuracy of inactive class predictions and including 8.3% of all off-target activities (Figure 5). In practice, selection of the ideal CI will be influenced by the relative proportions of on-target and off-target activities present in the given library of compounds to be assayed.

In our assay, the antianxiety class of culture-active compounds was least predictable (33.3% accuracy). This drug class also had the lowest number of example data sets (six) with which to build the statistical models, possibly contributing to the low prediction accuracy. Consistent with its misprediction, the average CI for the antianxiety class was low (Figure 1b). In spite of this low CI, an interesting misclassification trend was evident among the antianxiety drugs; three out of six were classified as anticonvulsant by the active drug class modeling (Figure 1a and b), or when the class was treated as an out-group (Figure 4). As benzodiazepines, each of these 'misclassified' anxiolytic drugs do, indeed, also possess clinical anticonvulsant activity.[12] Correspondingly, 100% of the anticonvulsants were classed as antianxiety when the anticonvulsants were treated as an unknown out-group. Three-dimensional depiction of class separation yielded by the top LDA biomarkers (Figure 2a) shows the close proximity of antianxiety and anticonvulsant drug classes in gene space (Figure 2b), consistent with the prediction of the functional overlap between members of these drug classes. Notably, this functional convergence was predicted by the system even though no explicit knowledge of shared efficacy was built into the model. A future direction will be to refine the ability of this system to identify cryptic clinical efficacy of existing drugs. Previously, we have shown that classes of therapeutic drug-induced cellular physiology can be modeled and used to predict the therapeutic action of drugs, even when no examples of the specific biochemical activity of the drug being assayed were used to build the predictive model. For example, selective serotonin reuptake inhibitors were correctly predicted as antidepressants even when only tricyclic and monoamine oxidase inhibitor antidepressants

were used for biomarker model construction.[11] Thus, it is possible to construct expression signatures of general therapeutic physiology that are potentially applicable to identifying drugs that act on previously unidentified targets, when those targets mediate physiology similar to a known drug class of interest. This potential to identify drug candidates in the absence of direct target knowledge faces the fundamental requirement of having some means of establishing a desirable cellular physiology beforehand, whether through drugs or other manipulations thought to have therapeutic relevance.

The approach we have described here could facilitate drug development in several respects. Distinguishing compounds with potential side effects as part of the same early step in which desirable activities are recognized can potentially diminish the pursuit of dead-end candidates. The use of stable mammalian cell line monocultures offers the prospect of lower inter-experimental variation and greater reliability than the use of primary cell cultures in this application. The capacity to assay individual drug candidates for multiple disparate activities in parallel can collapse into a single-step screens that would otherwise be conducted independently. In the future, it should be possible to expand the range of target efficacies to include a larger number of existing drug classes, as well as therapeutically relevant cellular physiologies that no small-molecule drugs currently mimic, such as can be induced by target gene suppression or growth factor treatment. The establishment of a quantitative multiplexed *in vitro* chemical genomic system for identifying and prioritizing novel activities provides a foundation for simplified and accurate gene expression-based drug discovery and development.

METHODS

Cell Culture

HCN-1A cells (ATCC) were grown to approximately 80% confluence in poly-L-lysine-coated 12-well cell culture plates (Corning) in DMEM (Gibco) supplemented with 10.0% FBS and pen/strep. Each drug was dissolved in DMSO (final DMSO concentration, 1.0 µl/ml) and used to treat a single culture for 24 h prior to harvest. Cultures treated with water-soluble drugs were supplemented with DMSO vehicle. Final drug concentrations were at pharmacologically relevant doses: alprazolam (1.0 µM), diazepam (10.0 µM), flurazepam (5.0 µM), triazolam (1.0 µM), zolpidem (10 µM), zopiclone (2.0 µM), carbamazepine (10.0 µM), ethosuximide (500.0 µM), felbamate (50.0 µM), gabapentin (40.0 µM), phenytoin (10.0 µM), riluzole (5.0 µM), valproate (200.0 µM), vigabatrin (100.0 µM), diltiazem (10.0 µM), nifedipine (10.0 µM), niguldipine (10.0 µM), nimodipine (10.0 µM), nitrendipine (5.0 µM), verapamil (10.0 µM), YS-035 (*N*-(-2-[3,4-dimethoxyphenyl]ethyl)-3,4-dimethoxy-*N*-methylbenzeneethanamine) (10.0 µM), aspirin (1.8 mM), dup.697 (7 µM), flurbiprofen (100.0 µm), ibuprofen (360.0 µM), indomethacin (8.0 µM), naproxen (25.0 µM), NS398 (10.0 µM), Sulindac (9.0 µM), DALDA (0.1 µM), DAMGO (0.1 µM), Dynorphin A (0.5 µM), Endomorphin (1.0 µM), Enkephalin (1.0 µM), U50488 (100.0 pM), U62066 (1.0 µM), lovastatin (1.0 µM), mevastatin (1.0 µM), pravastatin (1.0 µM), trazodone (2.0 µM), simvastatin (1.0 µM), X444241 (1.5 µM), clomipramine (1.2 µM), trimipramine (1.5 µM), amoxapine (2.0 µM), clozapine (2.5 µM), bupropion (2.5 µM), tranylcypromine (0.4 µM), fluvastatin (1.0 µM), citalopram (0.3 µM), desipramine (2.0 µM), maprotiline (1.0 µM), fluphenazine (0.08 µM), nortryptyline (0.7 µM), paroxetine (0.3 µM), 444242 (N-[[(4,5-dihydro-5-thioxo-1,3,4-thiadiazol-2-yl)amino]carbonyl]-L-phenylalanine) (11.0 µM), 444240 (α-[[[4,5-dihydro-5-thioxo-1,3,4-thiadiazol-2-yl)amino]carbonyl]amino]-((2-yridyl)piperazinyl)-(*S*)-benzenepropanamide) (0.15 µM), 444244 (OA-Hy*cis*-9-octadecenoyl-*N*-hydroxylamide oleoyl-*N*-hydroxylamide) (0.8 µM), 444264 (2-(hydroxamidocarbonylmethyl)-4-methylpentanoyl-homophenylalanine methylamide) (0.01 µM), 444251 (H-Cys[1]-Thr-Thr-His-Trp-Gly-Phe-Thr-Leu-Cys[10]-OH) (50.0 µM), 364205 (Galardin) (0.003 µM), 444249 ((2*R*)-[(4-biphenylylsulfonyl)amino]-*N*-hydroxy-3-phenylpropionamide) (0.07 µM), amitriptyline (1.0 µM), trifluprom azine (0.8 µM), phenelzine (0.8 µM), droperidol (0.16 µM), imipramine (2.0 µM), chlorprothixene (1.2 µM), chlorpromazine (0.9 µM), thiothixene (0.25 µM), perphenazine (0.33 µM), pimozide (0.05 µM), protriptyline (0.4 µM), risperidone (0.04 µM), thioridazine (4.0 µM), haloperidol (0.2 µM), doxepin (1.0 µM), trifluperazine (0.1 µM), loxapine (0.5 µM), fluoxetine (0.5 µM), fluvoxamine (1.5 µM), sertraline (1.4 µM), and iproniazid (0.6 µM). Drugs were purchased from Sigma, Tocris Cookson, or Calbiochem.

Sample Processing

Biotin-labeled cDNA was made with poly(T) primers from 15 µg of total RNA extracted from Trizol cell lysates. Gene expression was measured by hybridization of each sample (one sample per drug treatment) to the proprietary CuraChip microarray of ≈11 000 oligonucleotide probes. Slides were hybridized for 15 h with constant rotation at 30°C, washed for 30 min at room temperature (RT), incubated in streptavidin solution for 30 min at 4°C, washed three times for 15 min at RT, incubated in Cy3-conjugated detection buffer for 30 min at 4°C, and washed three times for 15 min at RT. Slides were scanned with a GMS 418 Scanner (Genetic Microsystems, Woburn, MA, USA) and analyzed with IMAGENE software (BioDiscovery, Marina Del Rey, CA, USA). Of ≈11 000 genes on the chip, 6047 were found to be expressed at least 3 × background.

Data Analysis

All genes detectable at least 3 × background after signal normalization were included in the 90 microarray data set for analysis. The modeling methods of LDA and SVM (linear kernel) were conducted either directly with ANOVA pre-filtered data or in an iterative process of selecting random inputs from pre-filtered feature space, a process we term SOGS. RF (125 trees, with 15 random inputs sampled at each set and a node size of 5) was conducted with ANOVA pre-filtered data. ANOVA data pre-filtering ($P<0.0001$) was accomplished within a leave-one-out cross-validation loop

to ensure that the resultant models were not biased by the marker selection process or over-fit within the confines of the experiment. The leave-one-out approach, where every sample data set was omitted from a discreet episode of model building, then presented to that model for classification, enabled all samples to be treated as unknowns for prediction by models naïve to the sample being classified. All models used equal weighting of class priors to prevent overemphasis of more heavily represented drug classes. To incorporate SOGS into LDA or SVM, two features are selected at random and evaluated by an index of class separation (Hotelling's T^2 in the case of LDA, and accuracy of the fitted model in the case of SVM), with the more predictive of the two features retained for model construction. This process is repeated until five features are collected, which are then used to predict the unknown. Bagging is accomplished by sampling with replacement from the training set. In all, 500 such model-building iterations are conducted, with the classification from each recorded. Unknown membership is assigned to the class receiving the most class assignments. The final accuracy of each method is assessed as the percent correct out of the total sample class predictions made.

The CI equals the ratio of the number of votes for the predicted class to the total number of votes cast across all classes. CI values were normalized to class number for model comparisons. All statistical algorithms were performed using the 'R' statistical software system (www.r-project.org).

DUALITY OF INTEREST

The work described was conducted by the authors while under the employ of CuraGen, Inc., which provided all financial support.

REFERENCES

1 Gasch AP, Huang M, Metzner S, Botstein D, Elledge SJ, Brown PO. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 2000; **11**: 4241–4257.
2 Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286**: 531–537.
3 Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD *et al.* Functional discovery via a compendium of expression profiles. *Cell* 2000; **102**: 109–126.
4 Schaner ME, Ross DT, Ciaravino G, Sorlie T, Troyanskaya O, Diehn M *et al.* Gene expression patterns in ovarian carcinomas. *Mol Biol Cell* 2003; **14**: 4376–4386.
5 Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001; **98**: 10869–10874.
6 Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME *et al.* Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002; **415**: 436–442.
7 Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000; **24**: 227–235.
8 van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; **415**: 530–536.
9 Johnson PH, Walker RP, Jones SW, Stephens K, Meurer J, Zajchowski DA *et al.* Multiplex gene expression analysis for high-throughput drug discovery: screening and analysis of compounds affecting genes overexpressed in cancer cells. *Mol Cancer Ther* 2002; **1**: 1293–1304.
10 Stegmaier K, Ross KN, Colavito SA, O'Malley S, Stockwell BR, Golub TR *et al.* Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nat Genet* 2004; **36**: 257–263.
11 Gunther EC, Stone DJ, Gerwien RW, Bento P, Heyes MP. Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles *in vitro*. *Proc Natl Acad Sci USA* 2003; **100**: 9608–9613.
12 Hardman JG, Limbird LE, Gilman AG. *Goodman and Gillman's The Pharmacological Basis of Therapeutics*, 10th edn. McGraw-Hill Companies, Inc.: New York, 2001.

Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles *in vitro*

Erik C. Gunther*†, David J. Stone*, Robert W. Gerwien, Patricia Bento, and Melvyn P. Heyes

CuraGen Corporation, 322 East Main Street, Branford, CT 06405

Edited by Floyd E. Bloom, The Scripps Research Institute, La Jolla, CA, and approved June 11, 2003 (received for review April 30, 2003)

Assays of drug action typically evaluate biochemical activity. However, accurately matching therapeutic efficacy with biochemical activity is a challenge. High-content cellular assays seek to bridge this gap by capturing broad information about the cellular physiology of drug action. Here, we present a method of predicting the general therapeutic classes into which various psychoactive drugs fall, based on high-content statistical categorization of gene expression profiles induced by these drugs. When we used the classification tree and random forest supervised classification algorithms to analyze microarray data, we derived general "efficacy profiles" of biomarker gene expression that correlate with antidepressant, antipsychotic and opioid drug action on primary human neurons *in vitro*. These profiles were used as predictive models to classify naïve *in vitro* drug treatments with 83.3% (random forest) and 88.9% (classification tree) accuracy. Thus, the detailed information contained in genomic expression data is sufficient to match the physiological effect of a novel drug at the cellular level with its clinical relevance. This capacity to identify therapeutic efficacy on the basis of gene expression signatures *in vitro* has potential utility in drug discovery and drug target validation.

pharmacogenomics | predictive efficacy | drug screening

Microarray-based gene expression patterns can be used as fingerprints of cellular physiology. The variety of cellular physiologies discernable by gene expression profile fingerprinting is expanding as an increasing range of cell types and cellular manipulations are investigated, and statistical methods of expression profile classification are refined. In yeast, distinctive profiles of genomic expression have been used to characterize cellular responses to diverse environmental transitions (1), functionally classify genetic manipulations, and discover a novel target for a drug of partially characterized function (2). In cancer studies, microarray data has been used to classify solid tumors (3), correlate tumor characteristics to clinical outcome (4), and cluster cell lines on the basis of their tissue of origin and response to drugs (5–9). In the area of toxicogenomics, large-scale gene expression analysis of toxin-treated cells and animals has yielded a highly accurate capacity to recognize the toxic potential of novel drug candidates (10–14), resulting in an increase in the efficiency of drug triage in the pharmaceutical development pipeline.

Multiple statistical methods have been applied to classification and recognition of expression profiles. Supervised classification analysis methods, which can classify patterns of novel data based on prior knowledge of sample classes, include linear discriminant analysis and genetic algorithm/K-nearest neighbors (11, 15), Fisher discriminant analysis (16), support vector machines (17), neural networks (18), and tree-based analysis (19). Here, we use human primary neurons treated with multiple members of multiple classes of antidepressant drugs, antipsychotic drugs, and opioid receptor agonists to generate DNA microarray gene expression data representative of these classes of treatment. We investigate whether example gene expression profiles from these drug treatments can be used to construct statistical models capable of predicting drug efficacy. We find that the classification tree (CT) and random forest (RF) supervised classification schemes can be used to predict the functional category of members of each of these drug classes with good accuracy, based on analysis of the gene expression profile induced by a drug.

Materials and Methods

Cell Cultures. Primary human neuronal precursor cells (Clonexpress, Gaithersburg, MD) were cultured for 7 days in growth media (GM) (50:50 DMEM/F12, 5% FBS, 10 ng/ml basic fibroblast growth factor, 10 ng/ml epidermal growth factor, 1:100 Clonexpress neuronal cell supplement, penicillin/streptomycin), and differentiated for 7 days in six-well plates at 900,000 cells per well in GM plus 100 μM dibutyrtyl cAMP, 20 ng/ml nerve growth factor, 1:100 matrigel, with 72-h media changes. Morphologically neuronal cells comprised ≈80% of the cultures.

Drug Treatments. Drugs were dissolved in DMSO and added to cultures at a final DMSO concentration of 0.04%. Drug concentrations represented pharmaceutically relevent doses: 2.0 μM amoxepine, 2.0 μM clomipramine, 2.0 μM desipramine, 1.0 μM doxepin, 2.0 μM imipramine, 1.0 μM maprotiline, 0.7 μM nortryptyline, 0.4 μM protriptyline, 1.5 μM trimipramine, 0.3 μM citalopram, 0.3 μM paroxetine, 1.4 μM sertraline, 0.4 μM tranylcypromine, 0.8 μM phenelzine, 0.6 μM iproniazid, 2.0 μM trazadone, 1.0 μM amitriptyline, 0.5 μM fluoxetine, 1.5 μM fluvoxamine, 2.3 μM bupropion, 1.0 μM chlorpromazine, 1.0 μM trifluperazine, 0.8 μM triflupromazine, 0.05 μM pimozide, 4.0 μM clozapine, 0.2 μM haloperidol, 0.04 μM risperidone, 0.5 μM loxapine, 0.1 nM BW373U86, 1.0 μM Enkephalin, 0.1 nM U50488, 1.0 μM U62066, 1.0 μM Endomorphin, 0.1 μM Tyr-D-Arg-Phe-Lys-NH2 (DALDA), 0.1 μM Tyr-D-Ala-Gly-NMe-Phe-Gly-ol (DAMGO), 0.1 μM Dynorphin A. Three control cultures were treated with DMSO only; two were treated with 5.0 μM phencyclidine (PCP) or 5.0 μM amphetamine. All treatments were 24 h in duration and conducted simultaneously. Drugs were purchased from Sigma or Tocris Cookson (Ellisville, MO).

Sample Processing. Cells were lysed in Trizol. Biotin-labeled cDNA was made by using 15 μg of total RNA with poly(T) primers. Gene expression was evaluated by hybridization to the proprietary CuraChip microarray (CuraGen, New Haven, CT) of ≈11,000 oligonucleotide probes. Slides were hybridized for 15 h at 30°C with constant rotation, washed for 30 min at room temperature (RT), incubated in streptavidin solution (4°C, 30 min), washed three times for 15 min at RT, incubated in Cy3-conjugated detection buffer (4°C, 30 min), and washed three times for 15 min at RT. Slides were scanned (GMS 418 Scanner, Genetic Microsystems, Woburn, MA) and analyzed by using

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CT, classification tree; RF, random forest; PCP, phenylcyclidine; SSRI, selective serotonin reuptake inhibitor.

*E.C.G. and D.J.S. contributed equally to this work.

†To whom correspondence should be addressed. E-mail: egunther@curagen.com.

www.pnas.org/cgi/doi/10.1073/pnas.1632587100

IMAGENE software (BioDiscovery, Marina Del Rey, CA). Of 11,000 genes on the microarray, ≈4,700 were found to be expressed at least 3× background.

Data Analysis. All genes detectable at least 3× background after signal normalization were included in the data sets for analysis. Data were prefiltered by using a generous Kruskal–Wallis filter ($P < 0.001$, ≈4,700 genes over 36 samples). Both CT and RF models were constructed by using the data sets from the 36 drug-treated samples, and calculated within a leave-one-out cross-validation loop to minimize the influence of marker prefiltering on model accuracy. The two highest-count markers selected by the CT decision tree were then removed from the data set and this process was repeated, then repeated again with the two top markers from the second iteration removed as well. For RF and each iteration of the CT algorithm, the samples were weighted such that an unknown would have an equal probability of falling within any class, and not default to the overrepresented class (antidepressants). The RF algorithm was also calculated with 1,000 trees grown and two random inputs attempted at each split. Two additional RF models were generated with the entire selective serotonin reuptake inhibitor (SSRI) or tricyclic class removed in a leave-one-subclass-out cross-validation loop. All statistical algorithms were performed by using the "R" statistical software system (www.cran.r-project.org).

Three-dimensional graphs were generated with DECISIONSITE (Spotfire, Somerville, MA). For CT iteration-one results, class separation provided by the four identified biomarkers was graphed for the four possible three-way combinations (leaving each gene out once). For RF results, class separation provided by the three markers with greatest confidence measures was graphed. The volumes occupied by the respective sample classes were delineated by lines interconnecting all of the correctly classified members in each group.

Results

Supervised Classification of Drug-Treated Samples. The supervised classification methods of RF and CT were used to analyze the gene expression profiles of drug-treated neurons. These methods have the advantage that examples of known classes can be used to build models of salient features that provide categorical distinction between the data sets used to build the models. These models can then be tested empirically with data sets of known class that were not used in model construction (cross validation). The ability of the model to correctly classify naïve samples serves as a measure of model quality. With both CT and RF methods, a "leave-one-out" training and testing series was conducted for all 36 drug-treated samples. Thus, 36 individual models were constructed, each trained with 35 example gene expression profiles, with one profile withheld from training for evaluation. After construction of each model, the profile excluded from the training set was tested by the model for assignment to one of the three drug treatment categories. Overall effectiveness of each method was calculated as the percent correct classifications out of the total 36 training and testing events conducted by the method.

Classification Tree. The CT method classified 32 of 36 expression profiles in the category corresponding to the therapeutic application of the drug used to treat the cells (88.9% correct classification) (Table 1). Of the 20 antidepressants, 18 were correctly classified. Of the eight antipsychotics, seven were correctly classified. Of the eight opioid receptor agonists, seven were correctly classified. Interestingly, four gene markers were sufficient to provide this level of resolution accuracy among these expression profiles, with pentaxin 3 (PTX3) and integrin-linked kinase (ILK) being sufficient to provide the majority of resolution between classes, as indicated by the high marker count for these genes (Table 2). Three-dimensional graphical representation of all possible three-way combinations of these four biomarkers illustrates the robust class separation provided by expression level comparison (Fig. 1).

Table 1. Classification of gene expression profiles induced by different drug treatments

Compound	True class	CT pred	RF pred	Subclass
Amoxepine	AD	AD	AD	Tricyclic
Clomipramine	AD	AD	AD	Tricyclic
Desipramine	AD	**OP**	AD	Tricyclic
Doxepin	AD	AD	AD	Tricyclic
Imipramine	AD	AD	AD	Tricyclic
Maprotiline	AD	AD	AD	Tricyclic
Nortryptyline	AD	AD	AD	Tricyclic
Protriptyline	AD	AD	AD	Tricyclic
Trimipramine	AD	AD	AD	Tricyclic
Amitriptyline	AD	AD	AD	Tricyclic
Citalopram	AD	AD	AD	SSRI
Paroxetine	AD	AD	AD	SSRI
Sertraline	AD	AD	AD	SSRI
Fluoxetine	AD	AD	AD	SSRI
Fluvoxamine	AD	AD	AD	SSRI
Tranylcypromine	AD	AD	AD	MAOI
Phenelzine	AD	**AP**	**AP**	MAOI
Iproniazid	AD	AD	**AP**	MAOI
Trazadone	AD	AD	AD	Atypical
Buproprion	AD	AD	AD	Atypical
Chlorpromazine	AP	AP	AP	Classic
Trifluperazine	AP	AP	AP	Classic
Triflupromazine	AP	AP	AP	Classic
Pimozide	AP	AP	AP	Classic
Clozapine	AP	**AD**	**AD**	Atypical
Haloperidol	AP	AP	AP	Atypical
Risperidone	AP	AP	AP	Atypical
Loxapine	AP	AP	AD	Atypical
BW373U86	OP	OP	**AD**	δ OPR
Enkephalin	OP	**AD**	**AD**	δ OPR
U50488	OP	OP	OP	κ OPR
U62066	OP	OP	OP	κ OPR
Dynorphin A	OP	OP	**AD**	κ/μ OPR
DALDA	OP	OP	OP	μ OPR
DAMGO	OP	OP	OP	μ OPR
Endomorphin	OP	OP	OP	μ OPR
Percent "correct"		88.9	83.3	

True class, known therapeutic utility; AD, antidepressant; AP, antipsychotic; OP, opioid receptor agonist. CT pred and RF pred, classes predicted by those methods for the expression profiles elicited by the drugs listed on the left; boldface designations, predicted therapeutic classes different from the "true" class. Subclass, the common pharmacological sub-classification; MAOI, monoamine oxidase inhibitor; δ, κ, μOPR, δ, κ, μ opioid receptor agonist, respectively.

Random Forest. The RF method classified 30 of 36 expression profiles in the category corresponding to the therapeutic application of the drug used to treat the cells (83.3% correct classification) (Table 1). Of the 20 antidepressants, 18 were correctly classified. Of the eight antipsychotics, seven were correctly classified. Of the eight opioid receptor agonists, five were correctly classified. The RF analysis identified 326 markers (not shown) used to construct the predictive models. The markers assumed an importance measure between zero and one. A large importance measure indicates that random permutation of that gene causes samples to be misclassified more often (hence

Table 2. Marker sets resulting from three sequential iterations of the CT analysis method

Marker identity	Marker count
CT iteration 1: 88.9% accuracy	
PTX3, pentaxin 3	35
ILK, integrin linked kinase	34
ENTPD6, ectonucleoside triphosphate diphosphohydrolase 6	1
GPCR CG50207	1
CT iteration 2: 72.2% accuracy	
SFRS7, splicing factor, arginine/serine-rich 7	34
ENTPD6, ectonucleoside triphosphate diphosphohydrolase 6	24
CBRC7TM_424 GPCR	1
APAF-1, apoptotic protease activating factor 1	1
ERMAP, erythroblast membrane-associated protein	8
CGFLC_31120	1
GPCR CG50207	1
LDHA, lactate dehydrogenase A	1
CT iteration 3: 80.6% accuracy	
LYPLA1, lysophospholipase 1	34
GPCR CG50207	26
CBRC7TM_424, GPCR	1
APAF-1, apoptotic protease activating factor 1	8
CGFLC_31120	1
LDHA, lactate dehydrogenase A	1

For each iteration, the two most frequent markers from the previous iteration were deleted from the expression data training set. Percent accuracy, the proportion of drug treatments correctly predicted by the respective marker set. G protein-coupled receptor (GPCR) CG50207 and CGFLC_31120 are novel sequences of GPCR and unknown function, respectively. Marker count, number of times out of the 36 model building episodes a particular gene was selected as a marker.

that gene is important). Thirty-two markers had an importance measure >0.35. Three markers had an importance measure >0.75: SFRS7 (splicing factor, arginine/serine-rich 7), SCG3 (secretogranin III), and hypothetical protein CG187232-01. Class separation provided by only these three top biomarkers (Fig. 2) is less distinct than the separation yielded by the biomarkers identified by the CT. This is probably because of the relative proficiency of the CT and RF algorithms with data sets containing a few strong markers, or a large number of weak markers, respectively.

Novel Subclass Prediction. The ability to accurately predict the functional class of an unrepresented drug type was tested by constructing models with an entire subclass omitted from the training data set. Two independent iterations of RF were conducted, with all expression data from SSRI-treated samples or tricyclic-treated samples withheld from the training data, respectively. Each model was then used to classify members of the antidepressant drug subclass withheld from training. The SSRI-naïve model correctly identified five of five (100%) SSRI treatments as antidepressants, and the tricyclic-naïve model correctly predicted 10 of 10 (100%) tricyclics as antidepressants, even though neither model was constructed with explicit examples of these respective subclasses.

Outgroup Identification by Graphical Depiction. A measure of model strength is the ability to distinguish inactive treatments from active drugs of various classes. In the most stringent case, a model should be able to exclude outgroup treatments from the target categories, even with outgroups not used to build the model. To determine whether the constructed models are capable of identifying novel outgroups as separate from established drug classes, we withheld vehicle-treated and nontarget class drug-treated control culture microarray data from the CT and RF training sets, to serve as unfamiliar outgroups. After the genes that distinguish the drug classes were identified, the top biomarkers from the CT and RF methods were depicted graphically, incorporating all of the data from each drug class as well as control samples (Figs. 1 and 2). A minimum volume encompassing the correctly classified samples was depicted by linearly connecting all members of each drug class. Vehicle-treated samples within these views were clearly excluded from the minimum volumes of each of the three drug classes, thereby illustrating a means by which outgroups may be effectively distinguished from therapeutic treatments, even when those outgroups are not included in the training data sets. PCP- and amphetamine-treated samples were effectively separated from the three main drug classes by RF, and fell very close to but outside the opioid category in the CT visualization.

Multiple Iterations of Classification Tree. Because relatively few genes were identified as drug class markers by the CT, we investigated the robustness of the approach in the absence of these markers by successively performing a second and third iteration of the analysis, excluding from the data set the predominant gene markers identified by the first and second iterations, respectively. The second iteration resulted in the identification of eight gene markers sufficient to provide 72.2% resolution accuracy. Misclassified as antidepressants in this second iteration were trifluperazine, clozapine, enkephalin, DALDA, and dynorphin. Misclassified as antipsychotics were clomipramine, phenelzine, iproniazid, and buproprion. Paroxetine was misclassified as opioid. Interestingly, the third iteration resulted in the identification of fewer gene markers (six) than the second iteration (Table 2), sufficient to provide greater resolution accuracy (80.6%) between treatment classes, even though the strongest markers from the second iteration were removed from the data set. Misclassfications made by the third iteration were identical to the second iteration, except paroxetine, buproprion and DALDA were correctly classified. Because CT and RF are based on similar recursive partitioning algorithms, the two approaches would be expected to identify many of the same markers. Indeed, both methodologies identified SFRS7 as a top marker. Overall, 55% of the markers from CT iterations 1–3, and six of the eight CT markers with counts >1, are among the 326 RF markers.

Discussion

The goal of the present study was to determine whether *in vitro* gene expression profiles of drug efficacy could be generated and used to predict the therapeutic classes of compounds administered to cell cultures. Recognition of toxin-induced expression profiles has proven possible because of the characteristic transcriptional responses to various types of cellular insult. Whether different classes of drugs applied at nontoxic doses elicit discernable signature expression profiles has not been investigated. In particular, we were interested in whether drugs distinguished by therapeutic indication at the patient level induce distinct expression profiles at the cellular level that correlate with clinical efficacy. To test this hypothesis, we selected three general classes of psychotropic compounds that might be expected to elicit transcriptional responses in human neurons: two that are used clinically (antidepressants and antipsychotics) and one that acts on a therapeutically relevant target class (opioid receptor agonists). Within each of these three categories, multiple subclasses were represented in order to capture general transcriptional features of each drug class.

CT and RF supervised classification schemes proved effective at predicting the physiological consequence of drug treatments

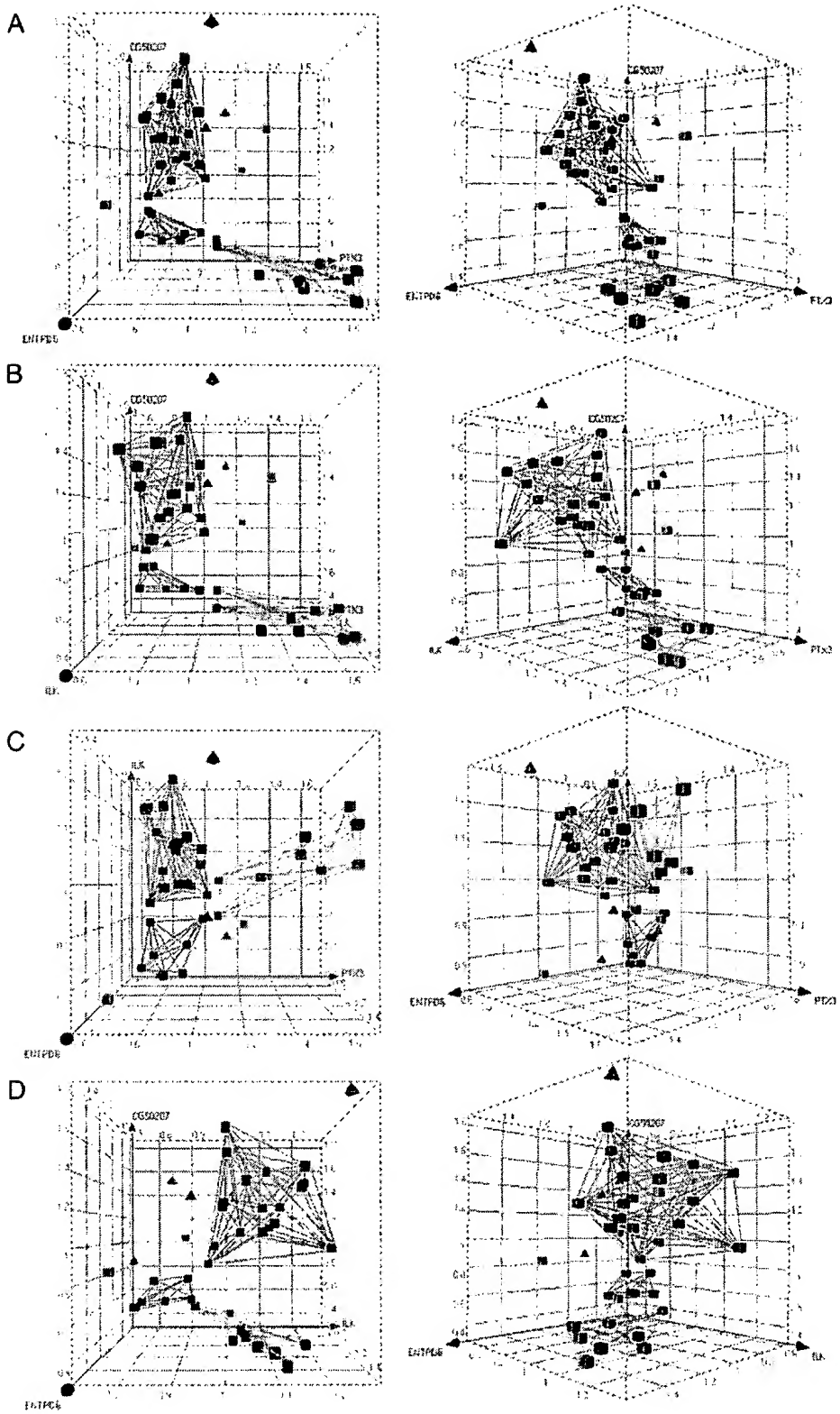


Fig. 1. Three-dimensional representation of class discrimination on the basis of biomarker expression: classification tree. All possible three-way combinations of the four-gene marker set from CT iteration-one are displayed: (CG50207 vs. ENTPD6 vs. PTX3) (*A*), (CG50207 vs. ILK vs. PTX3) (*B*), (ILK vs. ENTPD6 vs. PTX3) (*C*), and (CG50207 vs. ENTPD6 vs. ILK) (*D*). Axes represent relative expression levels of marker genes, with means set to 1.0. Each graph is shown perpendicular to the XY plane (*Left*), and from 45° rotation around the *y* axis (*Right*). Red, antidepressant; dark blue, antipsychotic; green, opioid; brown, PCP; black, amphetamine; light blue, vehicle control; squares, correctly predicted treatments; triangles, misclassified treatments. Lines connecting correctly predicted treatments delineate the volume occupied by each accurately defined sample class. Note distinct class separation and placement of untreated control samples outside the treatment classes.

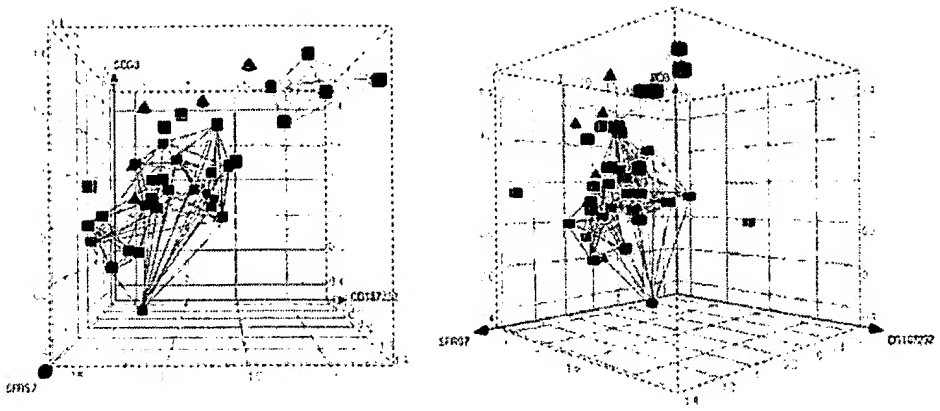


Fig. 2. Three-dimensional representation of class discrimination on the basis of biomarker expression: random forest. The three biomarkers with importance measures >0.75 are depicted: (SFRS7 vs. SCG3 vs. CG187232–01). Axes represent relative expression levels of marker genes, with means set to 1.0. The graph is shown perpendicular to the XY plane (*Left*), and from 45° rotation around the *y* axis (*Right*). Red, antidepressant; dark blue, antipsychotic; green, opioid; brown, PCP; black, amphetamine; light blue, vehicle control; squares, correctly predicted treatments; triangles, misclassified treatments.

on the basis of induced gene expression profile. The progressive leave-one-out strategy of model building and testing allowed us to comprehensively assess the ability of these schemes to accurately classify profiles induced by "novel" drugs (drugs not used to build the statistical predictive model). Because each of the three drug categories were weighted equally in our analyses, the accuracy of categorization based simply on random chance would be expected to be 33.3%. The 83.3% and 88.9% success rates of RF and CT, respectively, indicate that salient features of the distinct cellular physiologies induced by these drug classes are being identified by these methods, and that they can be used to correctly predict the clinical activity of unfamiliar compounds. In principle, this approach can be expanded beyond the three classes resolved here to encompass a large number of drug classes and cellular physiologies. Such high-content output is principally limited only by the number of pharmaceutical classes to which a given cell type is responsive.

The difference in accuracy between the CT and RF is likely caused by relative emphases on individual marker strength. Both methods partition feature-space to classify samples (20). CT performs this split once, to identify a few genes that each explain a large portion of the differences among classes. RF calculates many trees to identify many genes that each account for a small amount of the differences among classes. Although the CT method has several advantages over more traditional learning algorithms, the hierarchical nature of the data partition is especially sensitive to outliers, manifesting in high model variance and misclassification rates (20, 21). By virtue of its iterative character, RF improves model variance, generally yielding classification accuracies equal to or better than CT augmented by variance-reduction techniques such as boosting or bagging (20, 22, 23). Our finding that the first iteration of CT has a lower misclassification rate than RF (Table 1) suggests that these CT markers are especially important in distinguishing among the drug groups. This is supported by the subsequent two iterations of CT which each result in classification accuracies lower than the RF (Table 2).

How the specific biomarkers identified here relate systematically to the medical processes with which they correlate, and whether these biomarkers are also functional intervention points for further pharmaceutical development, are questions under investigation. Subclasses of drugs within each broad class are medicinally similar, but mechanistically heterogeneous [e.g., SSRIs, monoamine oxidase (MAO) inhibitors, and norepinephrine-uptake inhibitors act on distinct drug targets within multiple neurotransmitter systems, yet can exert similar effects on clinical depression]. This suggests that these drugs may share a common mechanism of action downstream from their proximal biochemical activities. Detected by our microarray were many genes involved in the transmitter systems targeted by the three drug classes, including MAO-a, MAO-b, multiple serotonin receptors, D2, and other dopamine receptors. Several expected targets, such as the serotonin transporter, norepinephrine transporter, and the δ, κ, and μ opioid receptors were not detected by the microarray, indicating they may be expressed below the level of detection, or that other, unconventional targets are the primary locus of action in this culture system. The presence of gene expression biomarkers that robustly predict drug class is consistent with the presence of convergent mechanisms of action among the respective subclasses of the three broad drug classes studied here.

Although the practical emphasis in expression profile fingerprinting to date has been to discern drug-toxicity profiles, drug-efficacy profiles may have utility in several important areas. One apparent application is to prioritize lead compounds early in drug development. The absence of a clear understanding of the mechanisms of psychosis and depression has prevented the development of a general cellular assay for drugs treating these disorders. For example, there is no one cellular assay for "antidepression." Apart from direct binding of specific drugs to different molecular targets, antidepressant activity can only be tested *in vivo* at the present time, presenting throughput limitations. Pharmaceutical development has thus been largely confined to "me-too" molecules of specific activity similar to existing compounds. Our discovery of efficacy profiles common to drugs of disparate activity but within a common therapeutic class provides a simple *in vitro* tool for finding functionally valuable drugs that act via a range of targets. Moreover, this method works whether the mechanism of action is known or novel. In our study, even when entire drug subclasses were left out of the RF training protocol, their medical utility was correctly identified upon testing of the models. The accurate classification of SSRI and tricyclic antidepressants by models naïve to these important pharmaceutical classes illustrates the prospect of discovering therapeutically valuable drugs that function by unknown mechanisms. Had SSRIs not been discovered before this study, for example, they could still have been accurately identified in our assay as having antidepressant activity.

In future refinements, it will be necessary to recognize false positives by including outgroups into the modeling protocol or establishing an alternative classification of "insufficiently close

to any efficacy profile." Three-dimensional graphical depiction of vehicle-treated control sample biomarker values placed them outside the minimum volume encompassing the respective drug classes (Figs. 1 and 2). Significantly, these control samples were omitted from the data sets used to construct the efficacy profiles. Thus, their effective partition supports the "insufficiently close to any efficacy profile" approach to novel outgroup identification. Control samples treated with compounds of other neuroactive classes (PCP and amphetamine) also fell distinctly outside the three drug class volumes in the RF depiction. In the CT depictions, although these neuroactive drugs fell technically outside the three classes, they were very close to the opioid treatments, perhaps because of shared aspects of cellular physiology. Any compound tested will necessarily have a position in n-dimensional biological descriptor space that falls closest to one of the centroids created by the different classes of drug treatment. In our present cheminformatic scheme, this closest centroid determines the class of the drug being tested. An optimized version of this assay may establish a minimum n-dimensional hypervolume surrounding all known members of a drug class, outside of which a given drug under evaluation may be classified as not belonging to any known therapeutic class. In practice, as new bioactive compounds are detected by the assay, the boundaries of each hypervolume may be empirically optimized by functionally assaying, *in vivo*, novel compounds that progressively define revised hypervolume borders.

Potentially of interest in analyses of this kind are trends in misclassification. For example, κ and μ opioid receptor agonist-induced expression profiles were predominantly correctly classified by both the RF and the first iteration of CT. However, δ opioid receptor agonist-induced profiles were predominantly classified as antidepressant profiles (Fig. 2). A question raised by this result is whether targeting the δ opioid receptor system would have utility in the treatment of depression. Studies of δ opioid agonist action in mouse models of depression, in fact, support the idea that this receptor system mediates depressive behavior (24). Thus, *in vitro* analyses of this kind may have utility in discovering novel therapeutic applications of drugs.

Another potential application of efficacy profiling is the validation of genes as therapeutic drug targets, an early step in the drug development process. Manipulation of a specific gene in culture, either by overexpression or gene silencing, may identify it as a drug target when a therapeutically relevant expression profile is induced. The derivation of biomarker profiles as signatures of specific pharmacological functions provides a foundation for *in vitro* assessment of the therapeutic relevance of presumptive disease-mediating biological molecules, and the efficacy of therapies devised to target them.

1. Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D. & Brown, P. O. (2000) *Mol. Biol. Cell* **11,** 4241–4257.
2. Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., *et al.* (2000) *Cell* **102,** 109–126.
3. Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 10869–10874.
4. Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., Kim, J. Y., Goumnerova, L. C., Black, P. M., Lau, C., *et al.* (2002) *Nature* **415,** 436–442.
5. Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C., *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9212–9217.
6. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature* **406,** 747–752.
7. Scherf, U., Ross, D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., Kohn, K. W., Reinhold, W. C., Myers, T. G., Andrews, D. T., *et al.* (2000) *Nat. Genet.* **24,** 236–244.
8. Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., van de Rijn, M., Waltham, M., *et al.* (2000) *Nat. Genet.* **24,** 227–235.
9. Califano, A., Stolovitzky, G. & Tu, Y. (2000) *Proc. Int. Conf. Intel. Syst. Mol. Biol.* **8,** 75–85.
10. Hamadeh, H. K., Bushel, P. R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J. C., Paules, R. S., *et al.* (2002) *Toxicol. Sci.* **67,** 232–240.
11. Hamadeh, H. K., Bushel, P. R., Jayadev, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J. C., *et al.* (2002) *Toxicol. Sci.* **67,** 219–231.
12. Waring, J. F., Ciurlionis, R., Jolly, R. A., Heindel, M. & Ulrich, R. G. (2001) *Toxicol. Lett.* **120,** 359–368.
13. Waring, J. F. & Halbert, D. N. (2002) *Curr. Opin. Mol. Ther.* **4,** 229–235.
14. Lakkis, M. M., DeCristofaro, M. F., Ahr, H. J. & Mansfield, T. A. (2002) *Exp. Rev. Mol. Diagn.* **2,** 337–345.
15. Mendez, M. A., Hodar, C., Vulpe, C., Gonzalez, M. & Cambiazo, V. (2002) *FEBS Lett.* **522,** 24–28.
16. Stephanopoulos, G., Hwang, D., Schmitt, W. A. & Misra, J. (2002) *Bioinformatics* **18,** 1054–1063.
17. Brown, M. P., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., Jr., & Haussler, D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 262–267.
18. Xu, Y., Selaru, F. M., Yin, J., Zou, T. T., Shustova, V., Mori, Y., Sato, F., Liu, T. C., Olaru, A., Wang, S., *et al.* (2002) *Cancer Res.* **62,** 3493–3497.
19. Zhang, H. & Yu, C. Y. (2002) *Front. Biosci.* **7,** c63–c67.
20. Hastie, T. & Tibshirani, R. (2001) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, Berlin).
21. Venables, W. N. & Ripley, B. D. (1999) *Modern Applied Statistics with S-PLUS* (Springer, Berlin).
22. Breiman, L. (2001) *Machine Learning* **45,** 5–32.
23. Amit, Y. & Geman, D. (1997) *Neural Comput.* **9,** 1545–1588.
24. Tejedor-Real, P., Mico, J. A., Smadja, C., Maldonado, R., Roques, B. P. & Gilbert-Rahola, J. (1998) *Eur. J. Pharmacol.* **354,** 1–7.

Microarray Analysis in Drug Discovery: An Uplifting View of Depression

Shawn E. Levy

(Published 28 October 2003)

Most drugs target a particular protein or cellular process: regulating, altering, or eliminating activity for the improved health of an individual. Good drugs are efficient and specific with strong action on a single biological pathway and minimal effects on other cellular processes. Drug target validation, identification of secondary drug effects, and the assessment of efficacy represent the most significant challenges in drug discovery and development. These challenges become even more significant if a well-defined cellular assay for the targeted disorder is unavailable. The lack of a clear understanding of the mechanisms of depression and psychosis has prevented the development of such a cellular assay, and psychoactive drug activity can only be tested in vivo at the present time. Thus, development of novel psychoactive drugs has been largely limited to generating compounds aimed at known therapeutic targets or with activities similar to existing drugs. The recent report by Gunther *et al.* describes the use of DNA microarrays to generate drug efficacy profiles and to match the physiological effect of a drug to its clinical relevance (*1*). The authors describe gene expression profiling experiments and analysis procedures to classify and predict the therapeutic class of a collection of psychoactive drugs belonging to three broad therapeutic classes. These experiments provide a technological and analytic foundation for the in vitro assessment of the therapeutic relevance of novel compounds with predicted psychoactive activities.

Every so often, a technology platform is developed that revolutionizes the scientific discovery process. High-throughput, automated DNA sequencing technologies have facilitated the sequencing and subsequent annotation of the genetic material from virus to yeast to four mammalian species, including humans and "man's best friend," with the recent report of the canine genome sequence (*2*). These sequencing efforts have provided a level of understanding and opportunity barely dreamed of 20 years ago. Around that time, the polymerase chain reaction (PCR) was first reported, and its potential to alter nearly every aspect of molecular biology was barely dawning. Genomic-scale profiling with microarrays was made possible by the genomic work that has followed the discovery of PCR, and this has substantially increased the resolution and efficiency with which cellular physiologies can be assayed at the molecular level by allowing the transcriptional state of every gene in a given genome to be profiled in a single experiment. Over the past 7 years, scientists the world over have embraced microarrays, with dozens of publications each month reporting a wide range of results generated by using microarrays in a multitude of experimental designs. The technical complexities associated with the production and use of microarrays, combined with the high-density data generated from genome-scale profiling experiments, has resulted in a renaissance of molecular biology to include traditionally disparate disciplines, such as computer science, engineering, theoretical mathematics, and statistics. The genome sequencing projects and genome-scale profiling technologies, such as expression microarrays, serial analysis of gene expression (SAGE), and others, have led to the evolution of entirely new research disciplines, such as Bioinformatics and Systems Biology. These technologies have also helped bring about an acceptance of research efforts aimed at generating and refining multiple hypotheses, rather than being limited by the strict testing of a single hypothesis. Microarrays evolved rapidly from their early beginnings—profiling less than 100 genes (*3*), through the whole-genome profiling of the diauxic shift in yeast (*4*), and their subsequent application to the study of human cancers (*5*). As the statistical methods associated with expression profiling have been refined, microarrays have proven to be particularly effective tools for class prediction and class discovery, most notably in the study of human cancers (*6-9*), including prediction of clinical outcome from the classification data (*6, 10, 11*).

Gunther *et al.* have applied a set of statistical tests to classify the gene expression profiles observed after psychoactive drug treatment of primary human neurons. The effectiveness of the statistical models to assign drug class accurately and to predict efficacy in the in vitro model was then evaluated. Using primary human neurons treated with three classes of psychoactive drugs (antidepressant, antipsychotic, and opioid receptor agonists), the authors analyzed an 11,000-gene oligonucleotide microarray to generate a training set of expression profiles that describes the transcriptional state specific to each of the 36 drugs. With the use of classification tree (CT) and random forest (RF) statistical techniques in a supervised approach, the authors were able to accurately predict the functional category of members of each of the drug classes on the basis of the gene expression profile induced by the drug treatment.

The CT technique is related to the hierarchical clustering methods that have become commonplace in the analysis of gene expression data. Classification trees are used to predict or explain the response of a categorical-dependent variable from their measurement of one or more predictor variables, which makes this analysis related to more traditional methods such as cluster and discriminant analysis. A simplistic example of how CTs work is sorting different balls by size. One effective way to classify them is to roll them down an incline that contains holes of increasing size. The first balls to fall would be classified as small, with medium and large balls separated as they proceeded farther down the incline according to their respective sizes. This simple example can be expanded and applied to various classification problems, including gene expression profiling. After the expression analysis of 36 different drugs representing three

Department of Biomedical Informatics, Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, 4th Floor EBL, 2209 Garland Avenue, Nashville, TN 37232-8340, USA. Telephone, 615-936-3000; fax, 615-936-3002; e-mail, shawn.levy@vanderbilt.edu

different classes (with each class having at least two subclasses based on biological action), the observed expression levels of the 11,000 genes present on the microarray were filtered to ~4700 well-measured genes across the 36 samples. The CT classification was performed over three iterations, where the two strongest class predictors from the first iteration were removed before the second. This was repeated for the third iteration, and each set of class prediction markers was used to classify the expression data in a leave-one-out validation scheme. The biological relevance and interrelatedness of the markers that were identified in the CT iterations are shown in Fig. 1, which indicates a diverse representation of many cellular processes. Out of the thousands of transcripts measured, only four genes were required to assign the profiled drugs to their correct classes with an accuracy of 88.9%. It is noteworthy that the third iteration of the CT (80.6% correct) was more accurate than the second iteration (72.2% correct), even though the four strongest classification markers from the previous two iterations were removed from the analysis. This could occur for several reasons, for example, genes involved in several pathways affected by a particular drug may be effective classification markers that are identified during different iterations of the CT algorithm. It may also be indicative of a weakness in the training set data. Indeed, because none of the iterations produced a completely accurate classification, the comprehensiveness of the microarray or its sensitivity for less abundant genes may not be sufficient to produce a perfect classification model. This should not limit the enthusiasm for the current study, however.

The results of analyzing the expression data set using the RF algorithm were similar to the results of the CT analysis with a slightly lower accuracy. The RF algorithm was developed by Leo Brieman and is a recursive approach similar to the CT method (see http://stat-http://www.berkeley.edu/users/breiman/rf.html for details on the RF approach), but it is conceptually quite radical. The RF method injects randomness into the use of tree-based statistical models, and although it is a "black box" process, it is more accurate and reliable than more standard tree-based models and does not over-fit the model. Generally speaking, the RF and CT methods differ in that CT methods base classifications on a small number of strong markers, whereas RF methods use small inputs from a larger number of markers. From a gene expression perspective, CT methods identify a few genes that can explain a large portion of differences between classes. The RF method calculates many trees that each account for small differences between the classes. Each method has relative advantages and disadvantages, and both were effective in distinguishing the three drug groups investigated by Gunther *et al.* Both methods were able to classify effectively a significant proportion of the 36 psychoactive drugs into three classes (antidepressant, antipsychotic, and opioid receptor agonists), without classifying vehicle controls or unrelated drug treatments that were included in the study. Although each of the classes contained drugs from multiple subclasses with different biological activities, the classification methods were able to group the subclasses together correctly using a leave-one-out cross-validation loop. How the specific biomarkers (Fig. 1) are related to the medical processes with which they correlate is the subject of further investigation. The diversity of biomarkers and the proper classification of the various subclasses suggest that each drug class may share a common mechanism of action downstream from the specific site of biological action that differs among the subclasses. It will be interesting to determine if the strongest markers identified in the Gunther *et al.* study represent potential targets for more effective therapeutic intervention or if they are secondary markers of an as-yet-unidentified process.

The use of microarray technologies in the field of pharmacogenomics is certainly not new. The first reports of drug target validation and identification of secondary drug effects by using microarrays to analyze drug effects in yeast followed closely on the heels of the first publications describing the profiling of the yeast transcriptome (*12*). From January to September 2003, there have been more than 100 publications describing the use of microarrays in pharmacogenomic or pharmacogenetic studies. The Gunther *et al.* study is unique and exciting in that it illustrates the potential application of genome-scale profiling data when robust statistical measures are applied to the data set to evaluate the complex question of drug efficacy. Building on this study, one can envision a growing collaborative data set from multiple cell lines and drug compounds that will provide a data scaffold on which to build a "master classifier" for drug action. This master classifier would be particularly valuable for the assessment of candidate and established compounds, as well as the refinement of novel statistical methods used to train the classifiers and to analyze the high-density expression profiles.

Although studies of this type are very interesting, there are important limitations to be considered. Drugs with complex action or multiple targets will be particularly difficult to analyze. Conditions where the primary sites of drug action are gene products essential for survival and in which drug effects that reach beyond a single cell type and have more complex effects in the whole organism will also be particularly challenging to evaluate in an in vitro model system. Finally, treated cells may show secondary effects or changes that provide strong markers for classification but are only distantly related to the primary drug effect. Of course, as the field evolves and matures and more studies are performed along the general design of the Gunther *et al.* study, the continued development and refinement of statistical methods and procedures will help to address these limitations. The early application of gene expression profiling to pharmacogenomics by Marton *et al.* used a yeast system to illustrate that expression profiling of drug-treated cells could be coupled with expression profiling of cells with gene deletions of the drug targets to mimic the inhibitory effects of the drug, which provides an evaluation of drug specificity and secondary effects that result in changes in gene expression (*13*). Updating this approach to current technologies, a drug evaluation scheme is possible that involves the classification of drug action by expression profiling, followed by the efficient and high-throughput assessment of the specificity of drug target action through the use of small interfering RNA (siRNA) to eliminate target gene expression in the cell model. Now that the use of siRNA in mammalian cells is more commonplace, this approach will certainly provide significant insight into possible secondary drug effects. Obvious limitations apply when essential genes are considered, but variations on this theme will undoubtedly add to the efficiency of drug target validation and screening.

The study by Gunther *et al.* illustrates the power of gene expression profiling and classification techniques in the evaluation and description of drug action. As the mathematical and statistical procedures to robustly analyze complex microarray data sets further expose the vast potential of genome-profiling

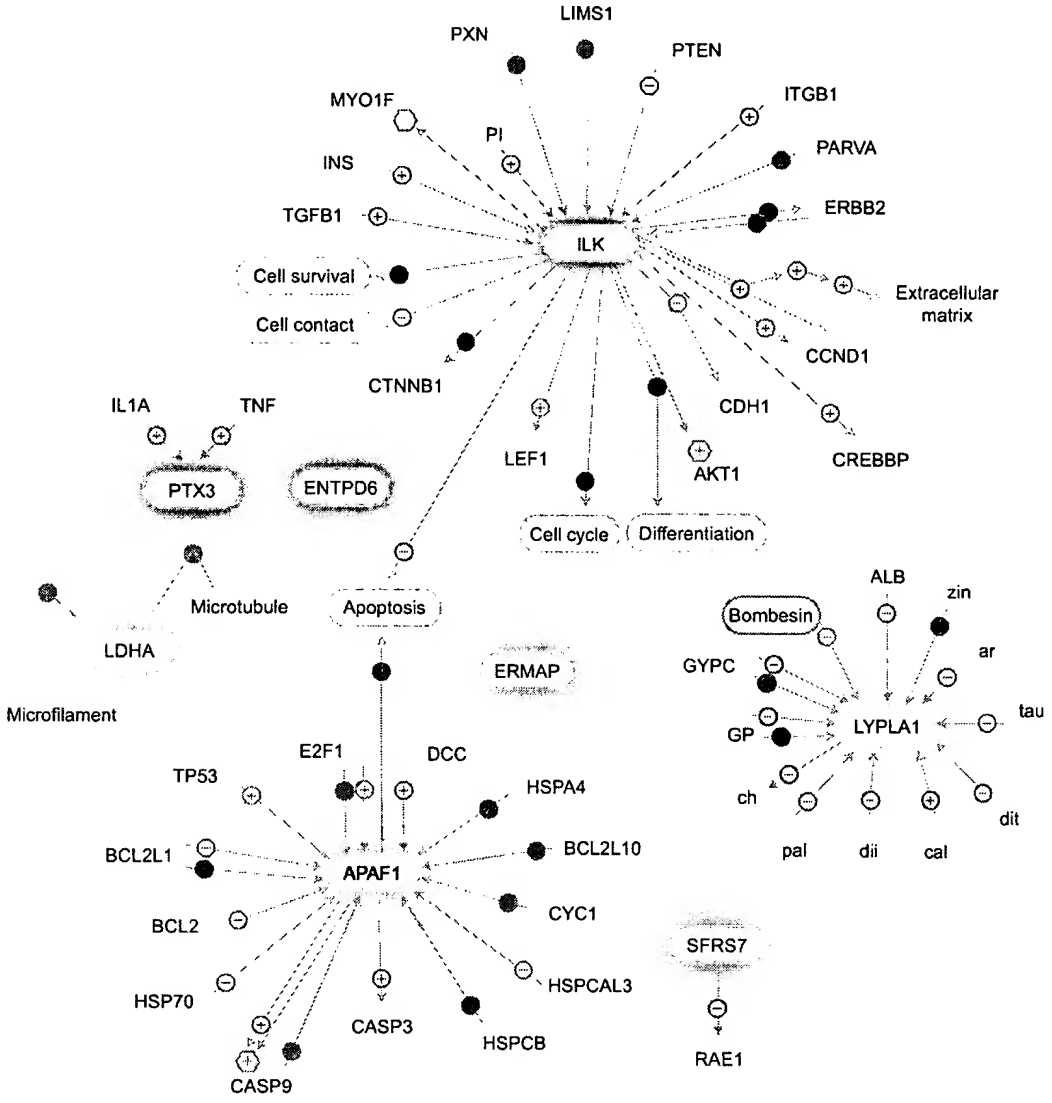


Fig. 1. The biological context of the key marker genes identified by the classification schemes used by Gunther *et al.* By using a natural language processing of all abstracts in PubMed along with protein-protein interaction data from the Biomolecular Interaction Network Database [BIND (http://www.bind.ca/)] and the Database of Interacting Proteins [DIP (http://dip.doe-mbi.ucla.edu/)], the biological context of the classification markers, the proteins and small molecules with which the markers interact, and the cellular processes in which the markers are involved can be efficiently represented graphically. The largest shapes represent proteins, small molecules, structures, or cellular processes. Proteins are shown in the white ovals and described by the human gene symbol (LocusLink). Small molecules are represented by white triangles; they are Ar, archidonoyl trifluoromethyl ketone; cal, calcium; ch, cholesterol; dii, diidopropylfluorophosphate; dit, dithiothreitol; GP, glycerophosphoethanolamine; pal, palmitoyl-CoA; PI, phosphatidylinositol(3,4,5)triphosphate; tau, taurocholate; and zin, zinc. Cellular processes are in yellow ovals. The small circles describe the relations between the large icons: Orange circles indicate a binding event (positive or negative regulatory effects caused by binding are indicated by a + or - sign in the circle, respectively). No sign indicates a binding event with an unknown or no regulatory effect, and blue circles indicate a regulatory event (+, positive regulation; -, negative regulation; and no sign indicates unknown or no regulatory effect). Protein modifications are indicated by three small orange hexagons on MYO1F (phosphorylation), AKT1 (phosphorylation), and CASP9 (self-cleavage). Similar to above, plus and minus signs indicate effects on regulation, if known. The ligand bombesin is represented by a purple oval. The strength of the marker gene to classify the drug treatment is highlighted with a halo. The strongest markers, which are those that represented genes with changes in expression in the first screen, are highlighted in blue. The other classification markers identified by subsequent iteration are highlighted in red and green, respectively. Markers that appeared in multiple groups are highlighted in the group where they had the most significance. All proteins not highlighted were added on the basis of their interactions with the marker genes.

data sets, the widespread implications for many disciplines begin to be appreciated. This is especially important in the field of pharmacogenomics, where any increase in the efficiency of drug target evaluation, toxicity, and efficacy can have dramatic effects in the discovery process. The most obvious application is the prioritization of lead candidate compounds early in development, with the process being effective for drugs with known or previously unknown mechanisms of action. Overall, the work of Gunther and colleagues is likely to have a positive impact on the application of gene expression profiling to drug discovery. The future looks especially bright for clinical conditions, such as depression or psychosis, where no cellular assay is able to define what constitutes an effective treatment. It is to be hoped that the application of genomic technologies will lead to a more complete understanding of the molecular basis of the underlying disease and the more efficient development of new therapeutics.

References

1. E. C. Gunther, D. J. Stone, R. W. Gerwien, P. Bento, M. P. Heyes, Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9608-9613 (2003).
2. E. F. Kirkness, V. Bafna, A. L. Halpern, S. Levy, K. Remington, D. B. Rusch, A. L. Delcher, M. Pop, W. Wang, C. M. Fraser, J. C. Venter, The dog genome: Survey sequencing and comparative analysis. *Science* **301**, 1898-1903 (2003).
3. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470 (1995).
4. J. L. DeRisi, V. R. Iyer, P. O. Brown, Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686 (1997).
5. J. L. DeRisi, L. Penland, P. O. Brown, M. L. Bittner, P. S. Meltzer, M. Ray, Y. Chen, Y. A. Su, J. M. Trent, Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**, 457-460 (1996).
6. A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson Jr., L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, L. M. Staudt, Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511 (2000).
7. T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, E. S. Lander, Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537 (1999).
8. M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts, V. Sondak, Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* **406**, 536-540 (2000).
9. C. M. Perou, T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown, D. Botstein, Molecular portraits of human breast tumours. *Nature* **406**, 747-752 (2000).
10. E. J. Yeoh, M. E. Ross, S. A. Shurtleff, W. K. Williams, D. Patel, R. Mahfouz, F. G. Behm, S. C. Raimondi, M. V. Relling, A. Patel, C. Cheng, D. Campana, D. Wilkins, X. Zhou, J. Li, H. Liu, C. H. Pui, W. E. Evans, C. Naeve, L. Wong, J. R. Downing, Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1**, 133-143 (2002).
11. S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. Kim, L. C. Goumnerova, P. M. Black, C. Lau, J. C. Allen, D. Zagzag, J. M. Olson, T. Curran, C. Wetmore, J. A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J. P. Mesirov, E. S. Lander, T. R. Golub, Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415**, 436-442 (2002).
12. M. J. Marton, J. L. DeRisi, H. A. Bennett, V. R. Iyer, M. R. Meyer, C. J. Roberts, R. Stoughton, J. Burchard, D. Slade, H. Dai, D. E. Bassett Jr., L. H. Hartwell, P. O. Brown, S. H. Friend, Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* **4**, 1293-1301 (1998).

Citation: S. E. Levy, Microarray analysis in drug discovery: An uplifting view of depression. *Sci. STKE* **2003**, pe46 (2003).